Essays on Market Microstructure and Pathwise Directional Derivatives

DISSERTATION

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)
im Fach Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
Frau Dipl.-Math. Jana Bielagk

Präsidentin der Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät:
Prof. Dr. Elmar Kulke

Gutachter:

1. Prof. Dr. Ulrich Horst (Humboldt-Universität zu Berlin)
2. Prof. Dr. Peter Imkeller (Humboldt-Universität zu Berlin)
3. Prof. Dr. Traian Pirvu (McMaster University, Kanada)

Tag der mündlichen Prüfung: 20. November 2017

Abstract

We analyze equilibrium problems arising from interacting markets and market participants, first competing markets with feedback and asymmetric information and then strategically interacting traders; moreover we develop and analyze a new notion of a pathwise directional derivative in the context of pathwise Malliavin calculus.

The first chapter analyzes a principal-agent game in which a monopolistic dealer (the principal) competes with a crossing network for trading with privately informed agents. The agents choose the best contract, i.e. quantity-price-combination, given their type and given the utility they could get from trading in the crossing network. The principal aims at designing pricing schedules that will maximize his profit. We analyze the structure of the principal's offer for different outside options ranging from getting nothing for not trading to outside options the principal will not be able to compete with. We formulate sufficient conditions for the existence and uniqueness of a solution to the dealer's problem and show that in our setting the introduction of the crossing network is beneficial for the agents. Additionally, we show that an equilibrium price for the feedback between dealer and crossing network exists and discuss its uniqueness.

In the second chapter we analyze the impact of performance concerns on a problem of equilibrium pricing. A derivative designed to make a non-financial risk hedgeable shall be priced such that supply equals the demand from a finite set of agents. The risk measure of each agent is specified by a Backward Stochastic Differential Equation (BSDE). The argument of the risk measure is the weighted sum of a risky position and the difference of the agent's trading gains and the average trading gains of all agents, punishing below average trading success and thereby requiring a strategic behavior of the agents. In spite of this strategic interaction, we are able to apply a representative agent approach (via weighted-dilated infimal convolution of the BSDEs' drivers) to obtain existence and uniqueness of the equilibrium market price of external risk. In the special case of entropic risk measures, we perform a parameter analysis, first analytically and then numerically, stating the influence of the different model parameters on the outcome. The equilibrium market price of risk is characterized by the solution to a quadratic BSDE.

The third chapter provides a link between classical and pathwise Malliavin calculus. On the one hand, there is the characterization of the Malliavin derivative via directional derivatives when the path varies with Cameron-Martin functions. On the other hand, there is the vertical derivative of Dupire, used by Cont and Fournié to establish a pathwise stochastic calculus of variations, where the variation is a certain step function. We propose a unifying approach by considering a wide range of variations for which we define directional derivatives: starting with Cameron-Martin functions we proceed with Hölder-continuous functions, continuous functions and discontinuous functions (regulated or of bounded variation). For the latter we rely on generalized Riemann-Stieltjes integrals – a novel approach within this field of research. Finally, we also introduce a notion of perturbation with a measure.

Zusammenfassung

Wir befassen uns mit Gleichgewichtsproblemen, die bei dem Zusammentreffen von Märkten und Marktteilnehmern entstehen, zuerst in einem Modell mit konkurrierenden Märkten mit Feedback und asymmetrischer Information und dann mit strategisch interagierenden Händlern. Zudem entwickeln und analysieren wir einen neuen Begriff der pfadweisen Richtungsableitung im Kontext des pfadweisen Malliavinkalküls.

Im ersten Kapitel analysieren wir ein Prinzipal-Agenten-Problem mit einem monopolistischen Dealer (dem profit-maximierenden Prinzipal), der mit einem Crossing-System um den Handel mit Agenten mit privater Information konkurriert. Die Agenten wählen entsprechend ihrem Typ und in Anbetracht ihrer Outside-Optionen die für sie beste Preis–Mengen–Kombination aus. Wir untersuchen die Angebote des Prinzipals in Hinblick auf ihre Struktur für unterschiedliche Outside-Optionen — von der einzigen Alternative des Nichthandelns und dafür nichts bekommen bis hin zu einer für den Prinzipal nicht zu überbietenden Outside-Option. Wir formulieren hinreichende Bedingungen für die Existenz und Eindeutigkeit einer Lösung des Maximierungsproblems des Dealers und zeigen, dass in unserem Modell die Einführung des Crossing-Systems für die Agenten vorteilhaft ist. Zudem zeigen wir, dass ein Gleichgewichtspreis für das Feedback-Problem existiert, und diskutieren die Frage nach dessen Eindeutigkeit.

Im zweiten Kapitel analysieren wir den Einfluss von vergleichender Leistungsbewertung von Händlern auf die Preisfindung im Marktgleichgewicht. Ein Derivat auf ein nicht-finanzielles Risiko soll den Preis bekommen, der am Markt der beteiligten Händler ein Gleichgewicht von Angebot und Nachfrage erzeugt. Das Risikomaß der Händler ist durch stochastische Rückwärtsgleichungen (BSDE) gegeben. Das Risiko eines Händlers setzt sich zusammen aus dem eigenen Risikoprofil, dem Erfolg des Handelns und dem durchschnittlichen Handelserfolg aller anderen Händler, so dass unterdurchschnittlicher Handelserfolg bestraft wird. Trotz des so entstehenden strategischen Handelns können wir einen repräsentativen Agenten bestimmen (mit Hilfe von gewichteter Konvolution der Treiber der BSDE) und mit dessen Hilfe die Existenz und Eindeutigkeit eines Gleichgewichtspreises für das nicht-finanzielle Risiko zeigen. Weiterhin können wir den Gleichgewichtspreis charakterisieren und im Spezialfall von entropischen Risikomaßen konkret berechnen. In diesem Spezialfall führen wir auch eine Parameteranalyse durch, um den Einfluss der verschiedenen Modellparameter auf das Marktgleichgewicht (Preis und Handelsmengen) zu bestimmen.

Das dritte Kapitel verknüpft klassischen und pfadweisen Malliavinkalkül. Malliavin-Ableitungen lassen sich einerseits mit Hilfe von Richtungsableitungen charakterisieren, wobei man den Pfad mit Cameron-Martin-Funktionen variieren lässt. Andererseits hat Dupire den Begriff der vertikalen Ableitung eingeführt, mit dessen Hilfe Cont und Fournié den pfadweisen stochastischen Kalkül entwickelten. Dieser Begriff erfordert, dass Pfade mit Treppenfunktionen variiert werden. Wir untersuchen verschiedene Klassen von Variationen, um beide Begriffe in einem Rahmen zu verbinden. Zuerst variieren wir Pfade mit Cameron-Martin-Funktionen, anschließend mit Hölder-stetigen Funktionen, mit stetigen Funktionen, mit unstetigen Funktionen beschränkter Variation und mit Regelfunktionen. Für die letzten beiden Fälle wenden wir, abweichend von der üblichen Vorgehensweise, das verallgemeinerte Riemann-Stieltjes-Integral an. Schließlich führen wir für unser Setting noch den Begriff der Variation mit einem Maß ein.

Contents

0 | Introduction

In the last decades, financial mathematics has motivated mathematical research in areas ranging from abstract probability theory to applied optimization and from game theory to the theory of market microstructure. This thesis contributes to different aspects of equilibrium theory with market imperfections and to pathwise stochastic analysis.

A central question at the intersection of market microstructure, mechanism design and pricing is how newly emerging alternative tradings systems and new financial products interact with existing markets and how prices and trade volumes are affected. The first two out of three chapters of this thesis are motivated by this question.

The first chapter originates from a project with U. Horst and S. Moreno-Bromberg, which resulted in the joint paper [BHMB16]. We analyze the competition between a monopolistic Dealer Market (DM) and a Crossing Network (CN), where the latter offers trades at a price derived from the former. The agents' types are private information, hence the dealer solves a screening problem to find the best offer, assuming a fixed price at the CN. In a second step, we analyze the feedback between DM and CN, leading to a (not necessarily unique) equilibrium price. We apply our theory to a setting with a Dark Pool (DP). The methods we use are mainly from mechanism design and variational analysis.

The second chapter originates from a project with G. dos Reis and A. Lionnet, which resulted in the joint paper [BLDR17]. Agents can trade in a financial product and a derivative written on a non-financial risk. Competition is not primarily between those products, but between the agents, which leads to strategic interaction of the agents and requires a game-theoretic approach. The goal is to set a price for the derivative such that the constant supply is met by the demand for the derivative. The combination of a Walrasian equilibrium problem with strategic interaction is solved by the representative agent method; mathematically we rely in particular on martingale representation under measure changes and Malliavin differentiation of BSDEs.

The third chapter, which is based on an ongoing cooperation with P. Imkeller, investigates pathwise directional (or variational) derivatives with the goal of providing a common setting for the vertical derivative from [Dup09] and the directional derivative for Cameron-Martin directions that is closely linked to the Malliavin derivative. While Malliavin calculus provides the motivation, the methods applied in this chapter are mostly from integration theory, including the less well-known gauge integral.

In what follows we will introduce the research problems of each chapter, describe our approach at solving these problems, present our main results and also review the related literature.

Introduction to Chapter 1

In traditional financial markets, liquidation of large positions (relative to the available liquidity) leads to an unfavorable price impact. In response to this problem, alternative trading venues such as Crossing Networks (CNs) have been established. Their most prominent features are that trading occurs at a price taken from a reference market (a stock market or Dealer Market (DM)), and execution of orders is uncertain. This leads to the question of how the prices and traded volumes in the DM are affected by the emergence of the CNs.

We model the price-setting market as a monopolistic DM run by a profit-maximizing principal (or dealer). The agents are traders who can choose between the incentive-compatible offer made by the dealer and their outside option, which is either not trading or trading in the CN, depending on which offers the highest (expected) utility. The price in the CN depends in a prespecified manner on the price schedule offered by the dealer, and simultaneously it determines the agents' outside option, thereby influencing the dealer's optimal strategy. This feedback loop leads to a fixed point problem, which we solve step by step.

First, we formulate the optimization problems of the dealer and the traders. The latter are parametrized by their types, which might represent e.g. inventory positions. The dealer does not know each agent's type, but only the distribution of types; hence, he faces a screening problem. We assume that whenever an agent is indifferent between DM and CN, the agent will accept the dealer's offer. For a given execution price in the CN, the dealer's objective is to maximize the profit from trading with the agents. We show that this constrained variational problem has a solution which is unique on the set of agents participating in the DM.

Second, we analyze the qualitative influence of the CN on the existing market. We prove that the set of reserved types, i.e. those who trade nothing in exchange for nothing, shrinks after the introduction of the CN; furthermore, for uniformly distributed types the spread narrows and the indirect utility, i.e. the highest attainable utility an agent can obtain from trading with the dealer, increases.

For the purpose of this comparison we first analyze the problem without a CN in the spirit of [BMR00]. Adding a non-trivial outside option complicates computations significantly, rendering an analysis by means of a maximum principle too cumbersome due to the discontinuities of the model. We overcome this difficulty by using an "accounting trick": Instead of excluding agents with whom trading is too costly we modify the dealer's cost function in such a manner that the costs equal the gains from trading with the previously excluded types of agents. Thus, these types can be regarded as non-excluded types giving zero revenue to the dealer.

We illustrate the results by means of several examples with and without CN. The structure of the sets of reserved types of agents, agents who are indifferent between DM and CN and agents being fully serviced by the dealer can become quite complex, as is demonstrated in particular in Example 1.4.9.

Having understood the dealer's problem, we show that for uniformly distributed types of agents there exists an equilibrium price, i.e., if sell and buy orders in the CN are executed at a specific price, the dealer will find it optimal to offer quantity-price bundles such that the best bid and best ask equal those from the CN. The proof is based on Tarski's fixed point theorem and the two results that there is a positive bid–ask spread (Lemma 1.5.1) and that types adjacent to reserved types are fully serviced (Lemma 1.4.7). We show in an example how the equilibrium price can be calculated numerically by recursively solving the dealer's problem, calculating the spread and solving the dealer's problem again for the new spread.

As an application of our model we consider agents aiming at liquidating their portfolios. They can choose to do so in a DM or in a Dark Pool (DP) and the agents' (uniformly distributed) types correspond to sizes of their portfolios. A DP allows investors to reduce their market impact by submitting liquidity that is shielded from the public view. The downside is that trade execution is uncertain: trades take place only when the matching liquidity is or becomes available, which is not known prior to the placing of orders. In case of a match, trades are typically settled at prices prevailing in an associated primary venue, such as the midpoint of the bid–ask spread. We obtain the existence of an equilibrium price and discuss the lack of the uniqueness thereof.

Related literature

The literature on the impact of market fragmentation and, more specifically, the impact of alternative trading venues on existing markets, has grown significantly in the last two decades, see for instance [GST$^+$13], [DVAW05] and [Ori12] for references to both theoretical and empirical papers.

Topics addressed in the market microstructure literature range from equilibrium models for competing markets (see e.g. [Glo94] and [PS03]) to the role of impatience and information and trading costs in the choice of markets (see e.g. [DDH13], [Ye16] and [DVAW09]) to the impact of emerging alternative trading venues on different aspects of market quality (see e.g. [Zhu14] and [BRW16]).

A common approach in the theoretical literature is to assume that the market participants trade only a single unit of the stock. For instance, in their seminal work, [HM00] derive conditions for the viability of the alternative trading institutions in a modeling framework where a random number of informed and liquidity traders, each buying or selling a single unit, chooses between a DM and a CN. In their model, dealers receive multiple single–unit orders and cannot distinguish between the informed and the liquidity orders.

In our model we have private information, but, contrary to [HM00], [BMR00], [Zhu14] and others, no public value. Therefore, there is no adverse selection and the monopolistic dealer cannot lose against informed traders.

The models considered by [Pag92], [Jul00] and [PSS$^+$08] feature this kind of screening problem. The former analyzes, in quite a general setting where the set of consumer types is a Polish space and the contract space an arbitrary compact metric space, the problem of a monopolist who faces both a screening problem (as in the work at hand) as well as a moral-hazard one relative to contract performance. [Jul00], on the other hand, only studies the screening problem in a finite-dimensional setting. This allows him to find a quasi-explicit representation of the optimal contract using Lagrange-multiplier techniques. He identifies conditions for the optimal contract to be separating, to be non-stochastic and to induce full participation. Furthermore, he also discusses the nature of the solution when bunching occurs. Neither of the two papers mentioned analyzes the case where the dealer's choices may have an impact on the structure of the reservation–utility function, which in turn would influence his decisions. Our study of such a feedback loop is novel and it is a crucial component in our analysis of the interactions between DMs and CNs, which is typically not unidirectional.

[Fag96] and [PSS$^+$08] study models of dealers competing for agents with private information about their types. The former analyzes a model with two equal dealers and an agent who may have one out of two possible types. The latter allows more than two agents and an unspecified type space, but still all dealers are identical. Contracts and the resulting competitive equilibria are called efficient if, among other requirements, the dealers obtain zero profit. Our setting differs significantly from both papers by having only one profit-maximizing dealer who competes with a CN whose price-schedule solely depends on the dealer's price, without any further maximization. While they have true Nash Equilibria, we have only a one-sided optimization. Without the symmetry from having identical dealers, it is not clear a priori that an equilibrium must exist in our setting.

As an application we think of DPs as a specific type of alternative trading venue. Simultaneous trading at a DM and a DP is modelled for instance in [HN14] and [KS15]. However, neither allows for an impact of off–exchange trading on the dynamics of the associated DM, which is precisely the feedback effect that we focus on.

Besides the cited theoretical works there exists a growing number of papers presenting empirical results on the effects of market fragmentation in general and dark trading venues in particular. For instance, [Bat97] empirically observes a positive effect of new OTC[1] trading (and hence increased market fragmentation) on NYSE-listed securities and a tightening spread. [Gre06] finds that risk-sharing benefits from CNs dominate fragmentation costs and cream-skimming[2]; if dealers are allowed to trade in the CN, then they can offer better prices. [NØ06] analyze different costs associated to trading in CNs such as direct trading costs (e.g. originating from the spread), adverse selection costs and opportunity costs from delayed trading. They find that implicit costs are larger than explicit trading costs. [NR14] present, besides a good literature review, evidence of informed traders in CNs, suggesting that information and price discovery happen in CNs due to concurrent trading. [DdJK15] empirically supports the theory of *cream-skimming* and finds a negative impact of dark trading on the related lit market. [AV15] finds empirical evidence for negative spillover effects of dark trading. [FP16] show that two-sided dark pools (i.e., dark limit order books (LOBs)) are beneficial, whereas the impact of one-sided dark pools (where crossing occurs e.g. at the midpoint of the bid–ask spread) is not clear and has an adverse-selection effect. We can sum up the empirical results with a statement from [GST$^+$13, Section 7.3]: "It is also possible that all types of dark pool trading activity may not have a uniform impact on the markets, given the different types of market structure that are clubbed in its definition."

[1]OTC (over-the-counter) – trading that occurs away from traditional exchange markets

[2]Cream-skimming refers to the effect that informed dealers tend to prefer the DM, whereas uninformed traders move to the DP, leading to a higher (adverse selection) risk for the dealer who faces the better-informed traders.

Introduction to Chapter 2

In this chapter we study the effects of social interaction between economic agents on a market equilibrium, on the efficiency of a securitization mechanism and on the overall risk. We consider a finite set $\mathbb{A}$ of N agents having access to an incomplete market consisting of an exogenously priced liquidly traded financial asset. The incompleteness stems from a non-tradable external risk factor, such as the amount of rain or the temperature, to which those agents are exposed. In an attempt to reduce the individual and overall market risks, a social planner introduces to the market a derivative written on the external risk source, allowing the agents in $\mathbb{A}$ to reduce their exposures by trading on it, and, under suitable conditions, completing the market.

Agents are price takers in a market with a stock and a derivative, they calculate how much they would like to buy or sell for a given price (θ) and while the stock market is exogenous, the market for the derivative clears due to what could be interpreted as a *Walrasian clearinghouse*[3]. This institution or person uses the knowledge about the agents' preferences in order to determine a price such that the market for the derivative clears. Then the actual trading will happen at this price. Completeness of the market is first assumed and then verified a posteriori in equilibrium. If demand for the derivative equals the supply and all shares of the derivative are held by agents in $\mathbb{A}$, then we say that we have *zero net supply*.

We add to this Walrasian equilibrium model some strategic behavior of the agents: Instead of just maximizing their profit from trading (or minimizing their risk), they also assess their trading gains relative to the average trading gains of all other agents. Each agent $a \in \mathbb{A}$ has an endowment H^a over the time period $[0, T]$ depending on both risk factors. Her investment strategy π^a in stock and the newly introduced derivative induces a gains process $(V_t^a(\pi^a))_{t\in[0,T]}$. For a given *performance concern rate* $\lambda^a \in [0, 1]$ the agent seeks to minimize the risk

$$\rho_0^a\bigg(H^a + \big(1-\lambda^a\big)V_T^a(\pi^a) + \lambda^a\Big(V_T^a(\pi^a) - \frac{1}{N-1}\sum_{b\in\mathbb{A}\setminus\{a\}} V_T^b(\pi^b)\Big)\bigg), \tag{0.0.1}$$

where ρ_t^a ($t \in [0, T]$) is a dynamic risk measure, which is derived from a BSDE with driver g^a and terminal condition given by minus the argument of ρ_0^a. The first two terms inside ρ_0^a correspond to the classical situation of an isolated agent a trading optimally in the market to profit from market movements and to hedge the financial risks inherent to H^a. The last term is the *relative performance concern* and corresponds to the difference between her own trading gains and the average trading gains achieved by her peers. Intuitively, as $\lambda^a \in [0, 1]$ increases, the agent is less concerned with the risks associated to her endowment H^a and more concerned with how she fares against the average performance of the other agents in $\mathbb{A}$. Thus, individual optimization is replaced by finding a Nash Equilibrium (NE).

Building upon the methods of equilibrium pricing from [HM07], [HPDR10] solve the model without strategic interaction by the representative agent approach[4], which is

[3]See for instance [Ath13, Chapter 1] for a conceptional exposition of the model of Arrow, Debreu and McKenzie and [Ath13, Section 2.2] on the Walrasian clearinghouse.

[4]For the origin of the representative agent approach see [Neg60], where aggregation of the agents into

known to be successful in providing Walrasian prices in this setting. With the addition of strategic interaction, and thus a combination of Nash and Walras equilibrium concepts, the success of the representative agent approach is not clear a priori. We do not only find the appropriate manner of aggregation such that the representative agent has preferences described in the same way as individual agents. We also verify that the representative agent approach provides a correct equilibrium price vector such that, given all agents adopt the (unique) NE strategy, the market for the derivative clears. An equilibrium in our setting is thus a collection of investment strategies and a market price of external risk.

We proceed as follows: At first, the performance concerns appear in the BSDE's terminal condition and not in the driver. For the purpose of finding the strategies that minimize the agents' risk functionals, it turns out to be more convenient to move the performance terms from the terminal condition into the driver and obtain a new BSDE. That BSDE can be minimized by minimizing the driver (cf. [HPDR10]). Suitable assumptions on the said driver (convexity, existence of unique minimizer) allow us to replace the expression depending on the other agents by an expressing that solely depends on the single agent's driver. As the terminal payoff is linear in the other agents' average performance, the NE for a given market price of risk can be computed by solving a linear system $A\pi = J$ (cf. [ET15, Section 3.2]). The matrix A contains the performance concern parameters $\lambda^{\cdot}$, while the right hand side vector J contains the solutions of N (independent) BSDEs and depends on θ.

The equilibrium market price of risk (θ^*) for zero net supply is characterized by the equation

$$0 = \sum_{a \in \mathbb{A}} \frac{\widetilde{Z}_t^{a,2}(\theta^*) - \mathcal{Z}^{a,2}(t, -\theta_t^*)}{(1 + \widetilde{\lambda}^a)}, \tag{0.0.2}$$

where $\widetilde{\lambda}^a = \frac{\lambda^a}{N-1}$, $\mathcal{Z}^a(t, \vartheta)$ is characterized by $\nabla_z g^a(t, \mathcal{Z}) = \vartheta$ and the pair $(\widetilde{Y}^a, \widetilde{Z}^a)$ solves the BSDE with terminal condition $\widetilde{Y}_T^a = -H^a$ and driver $\widetilde{G}^a(t, z^a) = g^a(t, \mathcal{Z}^a(t, -\theta_t)) + \langle \mathcal{Z}^a(t, -\theta_t), \theta_t \rangle - \langle z^a, \theta_t \rangle$.

For the special case of entropic risk, i.e. $g^a(z) = \frac{|z|^2}{2\gamma_a}$ (corresponding to expected exponential utility), Equation (0.0.2) can be rearranged and solved for θ^*. For more general drivers we show that, assuming that the market is complete, the representative agent approach gives θ^*. The representative agent's risk is given by a BSDE whose driver is a weighted infimal convolution of all agents' drivers. (0.0.2) suggests that the weights should be $\frac{1}{1+\widetilde{\lambda}^a}$ modulo some scaling. For the special case of entropic risk we verify that, under suitable assumptions, the market with the derivative is complete, we prove regularity results and we perform a parameter analysis. First, we calculate analytically the signs of the partial derivatives of aggregated risk with respect to the parameters λ^a $(a \in \mathbb{A})$, total and individual risk tolerance and number of units of the derivative (n), as well as the signs of the $\partial_n \theta_\cdot^R$ and $\partial_n B_\cdot^\theta$. Following the analytic parameter analysis we numerically perform a parameter analysis for the special case $\mathbb{A} = \{a, b\}$ on how initial trading activity π_0^a, the derivative's price B_0^θ and total and individual risk Y_0^w and Y_0^a $(a \in \mathbb{A})$ change as performance concerns and risk tolerance change.

---

a single economy and optimal Pareto risk sharing are equivalent to simultaneous individual optimization. Compare this with competitive equilibrium pricing, e.g. in [CHKP16].

Our main contribution is twofold: First, we demonstrate that the representative agent approach can be successfully applied to our model where Walrasian and Nash Equilibrium are combined. In particular, an a priori difficult system of BSDEs with interdependencies (cf. [ET15, Section 4.1]) is transformed into a single (solvable) BSDE, which can be interpreted as giving the representative agent's risk. Second, we obtain qualitative results concerning the effects of performance concerns when a derivative is introduced in a market in order to reduce exposure to previously non-hedgeable risks.

Related Literature

The theory of monetary, possibly convex, possibly coherent, risk measures was initiated by [ADEH99] and later extended by [FS02] and [FRG02]. One special class of risk measures, the so-called *g-conditional risk measures*, which are closely related to the so-called *g-conditional expectations* (see [Gia06]), are those defined through Backward Stochastic Differential Equations (BSDEs), see [Pen97], [ER09] and [BE09]. Our use of BSDEs is motivated by two general aspects. The first is that it generically allows to solve stochastic control problems away from the usual Markovian setup where one uses the Hamilton–Jacobi–Bellman approach in combination with theory of partial differential equations (PDEs), see e.g. [Tou13]. The second is that optimization can be carried out in closed sets of constraints without the assumption of convexity for which one usually uses duality theory, see [HIM05].

The question of the completeness of the resulting market has been addressed in some generality in the literature, and we refer for instance to [Sch17]. For questions about pricing and benefits of securities written on non-tradable assets, we refer in particular to [HM07] and [HPDR10], where the new derivative is priced such that supply equals demand, and more generally to [AR08]. For a good overview of equilibrium issues stemming the from market incompleteness as well as some solutions we point the reader to [KXŽ15] and references therein.

Our market model is based on that in [HM07] and [HPDR10][5]. While the former considers a problem of maximizing the expected terminal (exponential) utility from trading, the latter minimizes the risk corresponding to the solution of a BSDE with exponential utility only as special case. In both papers, an equilibrium price is such that the demand for the derivative (on the non-financial risk), given that all agents trade optimally, equals the constant supply. The representative agent approach is used in [HPDR10] whereas it is only briefly mentioned in [HM07, Remark 3.1]. In contrast to both papers, we do not start with the question of how to price the derivative in the given setting, but we take this method of pricing as given and analyze how the introduction of a strategic component (i.e., performance concerns) affects this method of pricing. This requires that we find a NE, prove its uniqueness before calculating the equilibrium price, and perform a parameter analysis with respect to (w.r.t.) the new parameters.

The importance of relative concerns in human behavior has been emphasized both in economic and sociological studies. In his pioneering monograph [Due49], Duesenberry does not only present arguments supporting "interdependence of preference systems", but he also explains data on saving which, without such interdependence, would be

[5] For more details on the differences between this work and [HPDR10] see Remark 2.6.7.

inconsistent with the theory. The utility functions representing such interdependencies, mostly developed and used only decades after Duesenberry's work, are said to exhibit a "keeping up with the Joneses" behavior. Contributions to this theory have been made, among others, by [Abe90], [Abe99], [Gal94], [CK02], [Góm07] and [XZ10]. A slightly different concern criterion uses the past consumption of the agent as a benchmark for the current consumption; [RH73] introduced this "habit formation" approach.

An application beyond private households and firms is presented for instance in [Ped15] and [HRAY10]: The social interaction component appears in the form of peer-based under-performance penalties known as "Minimum Return Guarantees"; the comparison is usually done via tracking a relevant market index, something quite standard in pension fund management.

Our performance functional compares total trading gains from a given time period with the average gains made by all other agents; it originates from [Esp10] and the papers referring to it, [ET15] and [FdR11].

[ET15] have a model of agents aiming to trade optimally in a market with given prices. Hence there is no Arrow-Debreu model of equating supply with demand and consequently there is no need for a representative agent. Furthermore, most of the work is done for exponential utility and not a general utility function. Only very briefly and solely for exponential utility do they consider a general equilibrium. They solve this problem without a representative agent in a manner similar to our Section 2.5.1. Their focus is not on trading strategies and pricing, but rather on the market index, i.e. the average wealth of all agents, and the impact of trading constraints.

On the other hand, [FdR11] focus on the role of investment constraints on the existence of equilibria. Again, there is no Arrow-Debreu type model, hence no representative agent. We solve unconstrained optimization problems and verify the admissibility a posteriori, where the admissibility criterion is deliberately chosen weak enough to not prevent existence of equilibria.

Introduction to Chapter 3

The motivation for this chapter comes from pathwise Malliavin calculus. Our study objects are functionals on the path space of a specific structure for which we want to obtain directional derivatives for different classes of directions (i.e., variations).

Stochastic calculus of variations, also known as Malliavin calculus, provides a link between stochastic and classical analysis. Since Malliavin's seminal papers in the 1970s, the topic has been developed and expanded in different directions and with different applications in mind, see e.g. [Bel06] for a brief historic overview, an exposition of different approaches and applications of Malliavin calculus.

A recent area of research develops pathwise notions of differentiation and integration in the context of the study of the regularity of Wiener functionals. A first approach to extend integration to paths of low regularity is Lyons' rough path theory (e.g. [LQ02], [FV10]), which introduced iterated integrals of rough paths. A further approach uses Fourier expansions and concepts of control between integrand and integrator, see e.g. [Gub04], [GIP15] and [GIP16]. A third approach addresses the representation theorem by Clark, Ocone and Haussmann: With the methodology from pathwise Malliavin calculus, the integrand can be explicitly given ([Nua06, Proposition 1.3.14] with conditional expectation or [CF13, Theorem 5.9] for a version with vertical derivative).

The notions of vertical and horizontal derivative in [Dup09] and [CF13] go beyond the directional derivatives from variation with Cameron-Martin functions, which are intimately connected with the ordinary Malliavin derivative. That connection, which is nicely presented e.g. in [Øk97], raises the question of whether or how to define directional derivatives beyond variations in the Cameron-Martin space in a manner that allows to maintain the link to the Malliavin derivative and obtain the vertical derivative[6] as special case.

We restrict our analysis to functionals that have a specific representation. This is not as strict an assumption as it might seem; the reason is that functionals on the path space can be factorized as functionals on the sequence space, see e.g. the first chapter of [Imk08] for details.

If a functional on $\Omega = C(I;\mathbb{R})$ for $I = [0,1]$ has the specific structure of a cylinder function, i.e., for a fixed orthonormal basis (ONB) of $L^2(I)$, $(h_k)_{k\in\mathbb{N}}$, one has $F(\omega) = f(\theta_1^h(\omega),\ldots,\theta_n^h(\omega))$ for $\theta_k^h(\omega) = \int_I h_k(t)\mathrm{d}\omega(t)$ $(k = 1,\ldots,n)$, then the Malliavin derivative of F with respect to time t evaluated for path ω can be calculated by the chain rule.

On the other hand, if for all Cameron-Martin functions $\gamma \in \mathcal{H}(I)$ there exists a function $\psi(t,\omega)$ satisfying

$$D_\gamma F(\omega) := \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left[F(\omega+\varepsilon\gamma) - F(\omega)\right] = \int_I \psi(t,\omega)\mathrm{d}\gamma(t),$$

then one may define $DF(t,\omega) := \psi(t,\omega)$ for $t \in I$ and $\omega \in \Omega$. As was pointed out in [Øk97], if the Malliavin derivative $DF(t,\omega)$ exists in both definitions, then both yield the same result.

While the above definition of $D_\gamma F(\omega)$ requires only perturbations of ω of the type $\omega+\varepsilon\gamma$ with $\varepsilon > 0$ and $\gamma \in \mathcal{H}(I)$ (Cameron-Martin space), the literature of pathwise calculus already features perturbations that go beyond variations with Cameron-Martin functions. For instance, the vertical perturbation in [Dup09] or [CF13] $\omega_T^h(s) := \omega(s) + h\mathbb{1}_{\{T\}}(s)$ $(s \in [0,T])$ or the variation with a step function in [SUV07] $\omega(s) + h\mathbb{1}_{[t,T]}(s)$ $(s \in [0,T])$.

Our goal is to find conditions on F for different classes of perturbations γ such that for fixed $\omega \in \Omega$

- the directional derivative $D_\gamma F(\omega) := \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left[F(\omega+\varepsilon\gamma) - F(\omega)\right]$ exists,
- the Malliavin derivative at ω, $DF(t,\omega)$, is integrable against γ in the Stieltjes sense, i.e., $\int_I DF(t,\omega)\mathrm{d}\gamma(t)$ is a well-defined Stieltjes integral,
- the equality $D_\gamma F(\omega) = \int_I DF(t,\omega)\mathrm{d}\gamma(t)$ holds.

[6] The horizontal derivative vanishes for functionals that do not explicitly depend on time.

We consider either cylinder functions or functions given by their infinite representation $F(\omega) = f((\theta_k^h(\omega))_{k\in\mathbb{N}})$; such an infinite representation is known to exist on the Wiener space. If we let $\beta_k(\omega) := \frac{\partial}{\partial x_k} f((\theta_k^h(\omega))_{k\in\mathbb{N}})$, then the Malliavin derivative of F, if it exists, has (by the chain rule) the form $DF(t,\omega) = \sum_{k=1}^{\infty} \beta_k(\omega)h_k(t)$. Similarly, the candidate for the directional derivative is $D_\gamma F(\omega) = \sum_{k=1}^{\infty} \beta_k(\omega) \int_I h_k(t)\mathrm{d}\gamma(t)$.

With F of the above specific structure, the three aspects of our goal give us four mathematical problems, namely finding conditions such that

- the Stieltjes integral $\int_I h_k(t)\mathrm{d}\gamma(t)$ exists in a given sense,
- the series defining $D_\gamma F(\omega)$ converges,
- $DF(\cdot,\omega)$ is integrable w.r.t. γ (in the same sense),
- integration and infinite summation (i.e., taking the limit to get the series) can be interchanged.

The classes of perturbations that we consider are Cameron-Martin functions, α-Hölder-continuous functions (vanishing at 0, for $\alpha \in (0,1)$), continuous functions (vanishing at 0), functions of bounded variation, regulated functions (i.e., functions which are uniform limits of step functions) and finite measures. Perturbing a path with a measure makes sense in our setting because all perturbations are, ultimately, integrators of Haar functions or of functions given by their orthogonal development w.r.t the Haar basis.

In the first part, we present the setting, introduce the notation and present the benchmark case of functions from the Cameron-Martin space. The first extension, by which we depart from the literature, is the case of Hölder-continuous functions, which calls for the application of the Young-Stieltjes (YS) integral. From this point onwards we use the Haar functions as ONB of $L^2(I)$ – this particular choice allows us to apply Ciesielski's isomorphism, which we also briefly recall. The next extension – from Hölder-continuous to merely continuous functions – goes slightly beyond Ciesielski's isomorphism and, more importantly, does not fall into the setting for Young's integration theory. While we do impose the rather strict requirement that $DF(\cdot,\omega)$ must have finite variation, we make no assumption on the p-variation of γ for any p. Instead, we notice that with a continuous integrator and a bounded integrand we are well in the setting of Riemann-Stieltjes (RS) integration.

The extension beyond continuous variation requires additional changes in the definition of our functional F: If we want to evaluate it for $\omega + \varepsilon\gamma$, that expression must fall into the domain of F. Consequently, we extend the domain of F and also space of functions that are allowed as perturbations of ω first to the set of functions of bounded variations and then to the set of regulated functions. The integrals go beyond both RS- and YS theory. Instead, we work with Lebesgue-Stieltjes (LS) integrals and the even broader class of generalized RS integrals (analogous to gauge integrals). Even though this is a true extension of the variations with continuous functions, the condition on the coefficients $\beta_\cdot(\omega)$ associated to F (or, more precisely, to f), remain the same. This condition (roughly speaking, it is $(2^{p/2}\beta_{pm}(\omega))_{p\in\mathbb{N}_0, m\in\{1,\dots,2^p\}} \in \ell^1$) is closely linked to the Haar functions, which increase with the factor $2^{p/2}$. We illustrate this link in Example 3.4.20 and Remark 3.4.21.

The last class of perturbations are measures. We present two possibilities of defining the corresponding directional derivatives. The first relies on the previous results by assigning to a measure the corresponding distribution function and thus returning to functions as variations. The second uses the abstract Lebesgue integral with respect to given measures.

The variations with $\mathbb{1}_{\{T\}}(s)$ $(s \in [0,T])$ and $\mathbb{1}_{[t,T]}(s)$ $(s \in [0,T])$ mentioned at the beginning are covered by our theory and emerge as a natural extension of variations with Cameron-Martin functions instead of a completely separate object.

Related literature

While Itô's stochastic calculus provides a stochastic integral that was not available in a pathwise (LS) sense, new developments in deterministic calculus (in particular in integration theory) permit in specific situations to return to a pathwise definition.

The pathwise approach to Malliavin and Itô calculus has received increasing attention in the last decade and a half. One motivation for this development is the need for results that are robust with respect to model uncertainty, as was the case for instance in [Vol16] or [PP16]. The former develops a pathwise integral from Föllmer's ideas (see [Föl81]), the latter proposes a controlled rough paths integral.

Another application is Monte Carlo Simulation, where pathwise derivatives are used to estimate derivatives, for instance in calculating Greeks in Mathematical Finance. See for instance [Gla04, Section 7.2].

A quick overview on functional Itô calculus can be found in [BCC16, Chapter 4]. For the link to [Föl81] we refer to the beginning of [BCC16, Chapter 5].

The theory of rough paths (see for instance [LCL07] or [FH14]), allows integration beyond semimartingale integrators (for instance integration w.r.t. fractional brownian motion, see [Nou12, Chapter 3]) and gives meaning to previously ill-posed stochastic PDEs.

Besides the applications mentioned above, perturbing paths in a non-smooth (in particular discontinuous) way is a common tool when working with Lévy processes, see e.g. [SUV07] where the variation is w.r.t. a step function $s \mapsto x \cdot \mathbb{1}_{[t,\infty)}(s)$ $(s \in I \subset \mathbb{R})$. So in some applications paths are not assumed to be continuous, but only càdlàg. We will start with continuous paths, but we will come across discontinuous paths and perturbations thereof from Section 3.6 onwards.

Notation

We introduce here only chapter-independent notation. Due to the differences between the topics of the chapters, we will introduce further specific notation in each of the chapters.

$\mathbb{R}$ shall denote the real numbers and $\mathbb{N} = \{1, 2, 3, \ldots\}$ the (positive) natural numbers. For closed intervals we use the notation $[a, b]$, whereas the corresponding open intervals are denoted (a, b). The left and right limits of of function f at x are denoted by $f(x_-)$ and $f(x_+)$, respectively. For $x \in \mathbb{R}$, $|x|$ denotes the Euclidean norm of x. For definitions we use the notation ":=". The subdifferential of a function f, evaluated at x, is denoted by $\partial f(x)$. If $A \subset \Omega$, i.e. A is a subset of the space Ω, then $A^c = \Omega \setminus A$ is the complement of A in Ω. If, for vector spaces A and B, we are given a function $f\colon A \to B$ and $x \in A$, then $f'(x)$ denotes the derivative of f at x. If time and space variables appear, we use $\dot{f}(t)$ to denote the derivative w.r.t. the time parameter. We write $f \equiv c$ if the function $f\colon A \to B$ is given by $f(x) = c$ for some $c \in B$ and all $x \in A$. The partial derivative of a function f w.r.t. variable x is denoted by $\frac{\partial f}{\partial x}$, $\nabla_x f$ or f_x for short if there is no danger of confusion with other subindices. The abbreviations $\mathbb{P} - a.a.$ and $\mathbb{P} - a.e.$ stand for $\mathbb{P}$-almost all and $\mathbb{P}$-almost everywhere, respectively, where $\mathbb{P}$ denotes a measure.

1 | Analysis of Crossing Networks Interacting with Dealer Markets via a Principal-Agent Model

1.1 Organization of this chapter

Our model and main results are presented in Section 1.2. Existence of a solution to the dealer's optimization problem is established in Section 1.3. Section 1.4 studies the impact of a CN on the spread. Section 1.5 establishes our result regarding the existence of equilibrium price schedules. A specific application to a portfolio-liquidation problem with DP trading is analyzed in Section 1.6; Section 1.7 concludes.

1.2 Model and main results

We consider a quote-driven market for an asset, in which a risk-neutral *dealer* engages a group of privately-informed *traders*[1]. The Dealer Market (DM) is described by a pricing schedule $T\colon \mathbb{R} \to \mathbb{R}$. In other words, q units of the asset are offered to be traded, on a take–it–or–leave–it basis, for the amount $T(q)$. For $q \in \mathbb{R}$, we refer to the pair $\big(q, T(q)\big)$ as a *contract*. We assume that $T(0) = 0$ and that T is absolutely continuous. Thus, we may write

$$T(q) = \int_0^q t(s)\mathrm{d}s, \quad q \geq 0,$$

and analogously for negative values of q. Here $t(s)$ is the marginal price at which the s-th unit is traded. As we shall see below, pricing schedules are, in general, not differentiable at zero. Hence, for a particular schedule T the *spread* is

$$\mathcal{S}(T) := |T'(0_+) - T'(0_-)| = |t(0_+) - t(0_-)|,$$

where $t(0_-)$ and $t(0_+)$ are the *best–bid* and *best–ask* prices, respectively. We denote by $C\colon \mathbb{R} \to \mathbb{R}$ the dealer's inventory or risk costs associated with a position q, e.g. the impact costs of unwinding a portfolio of size q in a limit order book. We assume that the mapping $q \mapsto C(q)$ is strictly convex, coercive[2] and that it satisfies $C(0) = 0$.

[1]Our dealer is called the *p*rincipal in the contract-theory jargon and the traders are usually referred to as the *a*gents.

[2]A function $C\colon \mathbb{R} \to \mathbb{R}$ is called coercive if $\lim_{|q|\to\infty} C(q) = +\infty$.

The traders' idiosyncratic characteristics are represented by the index θ that runs over a closed interval $\Theta := [\underline{\theta}, \overline{\theta}]$, called the set of *types*. We assume that zero belongs to the interior of Θ. Saying that a trader's type is θ means that if he trades q shares for $T(q)$ dollars his utility is $u(\theta, q) - T(q)$, where

$$u(\theta, q) := \theta\psi_1(q) + \psi_2(q).$$

We make the following assumption[3]:

Assumption 1.2.1. $\psi_1, \psi_2 \colon \mathbb{R} \to \mathbb{R}$ *are smooth functions that satisfy* $\psi_1(0) = \psi_2(0) = 0$, ψ_1 *is strictly increasing and* $C(q) - \psi_2(q) \geq 0$ *holds for all* $q \in \mathbb{R}$.

Thus far, with our choice of preferences the traders enjoy a type–independent *reservation utility* of zero, should they decide to abstain from trading in the DM. Such an action is commonly referred to agents choosing their *outside option*. As $C(0) = 0$, providing $\big(0, T(0)\big)$ is costless to the dealer and, since $\big(0, T(0)\big)$ yields all agents their reservation utility, in the absence of any other trading opportunity, we may equate the contract $(0, 0)$ to the traders' outside option.

Besides participating in the DM, each trader has the possibility to submit an order to a Crossing Network (CN) The latter is an alternative trading venue where trades take place at fixed bid/ask prices $\pi := \big(\pi_-, \pi_+\big)$, but where execution might not be guaranteed.[4] The possibility of trading in the CN modifies the traders' outside option to the extent that now they may choose between abstaining from all trading and earning zero or participating in the CN if the corresponding expected utility is non-negative. For a specific π, the quantity $u_0(\theta; \pi) \geq 0$ represents the expected utility of the θ-type investor who decides to take his (now extended) outside option. In a slight abuse of the language we also refer to $u_0(\cdot; \pi)$ as the agents' outside option(s). Following [DDH13] and [HM00] we focus on the case where a trader chooses exclusively between his outside option and trading in the DM, i.e., we do not allow for simultaneous participation in the DM and the CN. Initially we take π as given, but later we analyze the case where it is endogenously determined through the interaction between the DM and the CN. We work under the following assumption:

Assumption 1.2.2. *There is a fixed cost* $\kappa > 0$ *of accessing the outside option such that, for all* $\pi \in \mathbb{R}^2$, *the function* $u_0(\cdot; \pi)$ *can be written as* $u_0(\cdot; \pi) = \max\big\{\widetilde{u}_0(\cdot; \pi) - \kappa, 0\big\}$, *where* $\widetilde{u}_0(0; \pi) = 0$.

Trading in the DM is anonymous; the dealer is unable to determine a trader's type before he engages the latter. The only ex-ante information the dealer has is the distribution of the individual types over Θ, which is described by a density $f : \Theta \to \mathbb{R}_+$. In the sequel we specify the traders' and the dealer's optimization problems and analyze the impact of the CN on the DM, especially on its spread.

[3]Once an assumption has been made, we consider it to be standing for the remainder of the chapter.

[4]In other words, the CN presents agents with possibly better prices at the cost of an uncertain execution. CN trading often benefits agents who intend to unwind large positions, which might result in a price impact.

1.2.1 The traders' problem

Until further notice we consider π to be fixed. The problem of a trader of type θ is to determine, for a given pricing schedule T,

$$q_m(\theta) := \underset{q}{\operatorname{argmax}} \Big\{ u(\theta, q) - T(q) \Big\}$$

and then choose, for $q_m \in q_m(\theta)$, between his *indirect utility* $v(\theta) := u\big(\theta, q_m\big) - T\big(q_m\big)$ from trading in the DM and his outside option $u_0(\theta; \pi)$. As the supremum of affine functions, the indirect–utility function is convex.

The choice of a pricing schedule T induces a partition of the type space. We say that a trader of type θ *participates* in the DM if $v(\theta) \geq u_0(\theta; \pi)$, assuming that ties are broken in the dealer's favor. Conversely, we say that a trader of type θ *is excluded* from trading in the DM if $v(\theta) < u_0(\theta; \pi)$. For a given schedule T, we denote the set of excluded types by $\Theta_e(T; \pi)$. Observe that in the absence of a CN there is no loss of generality in assuming that all traders participate. We say that a trader of type θ is *fully serviced* if he earns strictly positive profits from interacting with the dealer.

1.2.2 The dealer's problem

According to the *Revelation Principle* (see e.g. [Mye91]) there is no loss of generality in focusing on direct-revelation mechanisms, i.e., those mechanisms where the set of types indexes the contracts. Furthermore, from the *Taxation Principle* (see e.g. [Roc85]) there is also no loss of generality in writing $\tau(\theta)$ instead of $T(q(\theta))$, where $\tau : \Theta \to \mathbb{R}$ is an absolutely continuous function. From this point on we shall, therefore, study our principal-agent game through books of the form $\big\{ \big(q(\theta), \tau(\theta)\big), \theta \in \Theta \big\}$ and drop T from the specification of the indirect–utility functions. We also write $\Theta_e(q, \tau; \pi)$ instead of $\Theta_e(T; \pi)$ for the set of excluded types.

At the onset, a trader of type θ could misrepresent his type by choosing a contract $\big(q(\widetilde{\theta}), \tau(\widetilde{\theta})\big)$, with $\widetilde{\theta} \neq \theta$. The dealer strives to avoid this situation, since he wants to exploit the information contained in the density of types. This requires that he offers *incentive-compatible* books, i.e. those that satisfy

$$\max_{\widetilde{\theta} \in \Theta} \big\{ u\big(\theta, q(\widetilde{\theta})\big) - \tau(\widetilde{\theta}) \big\} = u\big(\theta, q(\theta)\big) - \tau(\theta).$$

In the presence of an incentive-compatible book, the contract that yields a trader of type θ his indirect utility is precisely the one the dealer has designed for him.

Since the dealer is risk neutral, his goal is to maximize his expected income from engaging the traders. Taking into account the impact of the CN on the traders' optimal actions, his problem is to devise (q^*, τ^*) so as to solve the problem

$$\mathcal{P}(\pi) := \begin{cases} \sup_{(q,\tau)} \int_{\Theta_e^c(q,\tau;\pi)} \Big(\tau(\theta) - C\big(q(\theta)\big)\Big) f(\theta) \mathrm{d}\theta \\ \text{s.t.} \\ (q, \tau) \text{ is incentive-compatible}, \\ \tau \text{ is absolutely continuous}. \end{cases}$$

Due to the *Envelope Theorem*, if a contract $\{(q(\theta), \tau(\theta)), \theta \in \Theta\}$ is incentive-compatible, then $\psi_1(q(\theta))$ belongs to the subdifferential $\partial v(\theta)$. Since for almost all $\theta \in \Theta$ it holds that $\partial v(\theta) = v'(\theta)$ and ψ_1 is strictly increasing, we have for almost all $\theta \in \Theta$ that

$$q(\theta) = \psi_1^{-1}\big(v'(\theta)\big). \tag{1.2.1}$$

Therefore, starting from a convex indirect–utility function we can recover, for almost all types, the quantities in the incentive-compatible book that generated it. Furthermore, the indirect–utility function may be written as

$$\begin{aligned} v(\theta) &= \theta\psi_1\big(\psi_1^{-1}\big(v'(\theta)\big)\big) + \psi_2\big(\psi_1^{-1}\big(v'(\theta)\big)\big) - \tau(\theta) \\ &= \theta\, v'(\theta) + (\psi_2 \circ \psi_1^{-1})(v'(\theta)) - \tau(\theta). \end{aligned} \tag{1.2.2}$$

It follows from Equations (1.2.1) and (1.2.2) that the traders' indirect–utility function contains all the information about the quantities and the pricing schedule, which allows us to write $\Theta_e^c(v;\pi)$ instead of $\Theta_e^c(q,\tau;\pi)$. In particular, introducing the functions

$$\widetilde{K}(q) := C\big(\psi_1^{-1}(q)\big) - \psi_2\big(\psi_1^{-1}(q)\big) \quad \text{and} \quad i(\theta, v, q) := \theta q - v - \widetilde{K}(q)$$

and denoting by $\mathcal{C}$ the cone of all real-valued convex functions over Θ, we can restate the dealer's problem as

$$\mathcal{P}(\pi) = \sup_{v\in\mathcal{C}} \int_{\Theta_e^c(v;\pi)} i\big(\theta, v(\theta), v'(\theta)\big) f(\theta) \mathrm{d}\theta.$$

We prove in Theorem 1.2.5 below that, under suitable assumptions, Problem $\mathcal{P}(\pi)$ admits a solution. The latter is, in fact, quasi-unique in the sense that on the set of participating types the solution is indeed unique. However, agents are excluded by offering them any incentive-compatible indirect–utility function that lies below u_0. In other words, there is no uniqueness on the set of excluded types. From the agents' point of view there is no ambiguity: they either trade with the dealer or they take their outside option. The non-uniqueness is also no issue for the dealer, since it only appears in subdomains of the type space that he does not access. With this in mind, in the sequel we denote by $v(\cdot;\pi)$ "the" solution to Problem $\mathcal{P}(\pi)$.

Assumption 1.2.3. *The functions ψ_1, ψ_2 and C are such that $\widetilde{K}$ is strictly convex, coercive, continuously differentiable and it satisfies $\widetilde{K}'(0) = 0$.*

Determining the set of types who do participate but who get zero utility is essential to our analysis, since it is precisely at the *boundary types* where the best–bid and best–ask prices $t(0_-)$ and $t(0_+)$ are determined. We prove in Lemma 1.4.7 that, by virtue of Assumption 1.2.2, these limits are always well defined. For any $v \in \mathcal{C}$, we shall refer to

$$\Theta_0(v) := \{\theta \in \Theta \mid \ v(\theta) = 0\}$$

as the set of *reserved traders*. Whenever we refer to the reserved set corresponding to the solution $v(\cdot;\pi)$ to $\mathcal{P}(\pi)$ we write $\Theta_0(\pi)$. We prove in Proposition 1.3.2 that there is no loss of generality in assuming that any admissible $v \in \mathcal{C}$ satisfies $v(0) = 0$; thus, $\Theta_0(v) \neq \emptyset$.

Remark 1.2.4. *A well defined spread requires $\Theta_0(\pi)$ to be a proper interval $[\underline{\theta}_0(\pi), \overline{\theta}_0(\pi)]$, which will follow from Assumption 1.2.2, and that there exists $\epsilon > 0$ such that $(\underline{\theta}_0(\pi) - \epsilon, \underline{\theta}_0(\pi))$ and $(\overline{\theta}_0(\pi), \overline{\theta}_0(\pi) + \epsilon)$ belong to the set of fully-serviced traders. The existence of such an ϵ is proved in Lemma 1.4.7. Economically, this conditions means that the CN is not beneficial for low-type traders. We shall encounter several instances where the proofs of our results concern conditions on points to the left of $\underline{\theta}_0(\pi)$ or to the right of $\overline{\theta}_0(\pi)$ that are analogous. So as to streamline the said proofs, whenever we find ourselves in one of these "either–or" situations, we deal only with the positive types.*

We are now ready to state the first main result, whose proof is given in Section 1.3 below.

Theorem 1.2.5. *Problem $\mathcal{P}(\pi)$ admits a solution, which is unique on the set of participating types.*

Our second main result concerns the effect of the CN on the spread and the set of participating traders if, disregarding negative expected unwinding costs, the dealer can match the CN.

Assumption 1.2.6. *There exists an incentive-compatible book $\big\{(q_c(\theta), \tau_c(\theta)), \theta \in \Theta\big\}$ such that for almost all $\theta \in \Theta$ it holds that $u\big(\theta, q_c(\theta)\big) - \tau_c(\theta) = u_0(\theta; \pi)$.*

Assumption 1.2.6 implies that $u_0(\cdot; \pi)$ is also a convex function. The case where $u_0(\cdot; \pi)$ is concave is somewhat simpler, since it boils down to exclusion without matching.

The following theorem analyzes the impact of the CN on the DM and the traders' welfare.

Theorem 1.2.7. *For a given price $\pi = (\pi_-, \pi_+)$ let $\mathcal{S}_m$ and $\mathcal{S}_o$ be the spreads with and without the presence of the CN and $v(\cdot; \pi)$ and v_o the corresponding indirect–utility functions, respectively. In the presence of the CN*

1. *less types are reserved, i.e. $\Theta_0(v_o) \supseteq \Theta_0(\pi)$. Furthermore, the inclusion is strict if there exists $\theta \in \Theta$ such that $u_0(\theta; \pi) > v_o(\theta)$;*

2. *if the types are uniformly distributed ($f = (\overline{\theta} - \underline{\theta})^{-1} \mathbb{1}_\Theta$), the spread narrows, i.e. $\mathcal{S}_o \geq \mathcal{S}_m$;*

3. *the typewise welfare increases, i.e. $v_o(\theta) \leq v(\theta; \pi)$ for all $\theta \in \Theta$.*

In the sequel we use the subindices "m" and "o" to distinguish structures or quantities with and without a CN, respectively.

1.2.3 Equilibrium

It is natural to assume that pricing in the DM has an impact on the pricing schedule π. For example, the CN could be a Dark Pool (DP), where trading takes place at the best bid and best ask prices of the primary market. We analyze such an example, within a portfolio-liquidation framework, in Section 1.6. The pecuniary interaction between the DM and the CN, however, is not unidirectional if the dealer anticipates the effect that his choice of book structure has on the CN. Our main focus is the impact of the CN on the spread in the DM. Specifically, if we denote by $t(0;\pi) := \big(t(0_-;\pi), t(0_+;\pi)\big)$ the best–bid and best–ask price in the DM for a given CN price schedule π, then we call π^* an *equilibrium price* if $\pi^* = t(0;\pi^*)$.

We make the following natural assumption on the impact of π on the traders' outside option.

Assumption 1.2.8. *Let $\pi_1 \leq \pi_2$, where "$\leq$" is the lexicographic order in $\mathbb{R}^2$; then for all $\theta \in \Theta$ it holds that $u_0(\theta;\pi_1) \geq u_0(\theta;\pi_2)$. Furthermore, we assume that there exists $\big(\underline{\pi}_-, \overline{\pi}_+\big) \in \mathbb{R}^2$ such that $u_0(\cdot;\pi) \leq 0$ for all (π_-, π_+) such that $\pi_- \leq \underline{\pi}_-$ and $\overline{\pi}_+ \leq \pi_+$.*

The following is our main result on the existence of an equilibrium price.

Theorem 1.2.9. *If types are uniformly distributed, then the mapping $\pi \mapsto t(0;\pi)$ has a fixed point.*

Summarizing, we have that the dealer can correctly anticipate the movements in prices in the CN when he designs the optimal pricing schedule for the DM. Furthermore, the presence of the CN is beneficial in terms of liquidity, market participation and the traders' welfare.

Remark 1.2.10. *The uniformity of the distribution of types in Theorems 1.2.7 and 1.2.9 can be relaxed, which is something we postpone to the corresponding proofs, where the required notation is introduced.*

1.3 Existence of a solution to Problem $\mathcal{P}(\pi)$

In this section we prove the existence of a solution to the dealer's problem in the presence of a CN. Even though, strictly speaking, this result is a particular case of Theorem 4.4 in [Pag92], for the reader's convenience we present a proof in our simpler setting. Some of the arguments are somewhat standard, but we include them for completeness. The first important result that we require is that the dealer's optimal choices will lead to him never losing money on types that participate.

Proposition 1.3.1. *If $(q^*, \tau^*) \colon \Theta \to \mathbb{R}^2$ is an optimal allocation, then for all participating types it holds that $\tau^*(\theta) - C\big(q^*(\theta)\big) \geq 0$.*

Proof. Assume the contrary, i.e. that the set

$$\widetilde{\Theta} := \left\{ \theta \,\middle|\, v(\theta;\pi) \geq u_0(\theta;\pi), \tau^*(\theta) < C\big(q^*(\theta)\big) \right\},$$

where $v(\theta;\pi) = u\big(\theta, q^*(\theta)\big) - \tau^*(\theta)$, has positive measure. Define a new pricing schedule via

$$\widetilde{\tau}(\theta) := \max\left\{\tau^*(\theta), C\big(q^*(\theta)\big)\right\}.$$

The incentives for types in $\widetilde{\Theta}^c$ do not change, since their prices remain unchanged, whereas prices for others have increased. Agents who did not participate still don't participate after the change. Profits corresponding to trading with types in $\widetilde{\Theta}$ increase to zero. As a consequence the dealer's welfare strictly increases, which violates the optimality of (q^*, τ^*). $\square$

A consequence of Proposition 1.3.1 is that, together with Assumption 1.2.3, it allows us to restrict the admissible set of the dealer's problem (so far it is $\mathcal{C}$) to a compact one. We prove this in several steps.

Lemma 1.3.2. *If $v\colon \Theta \to \mathbb{R}$ is a non-negative convex function that solves $\mathcal{P}(\pi)$, then $v(0) = 0$.*

Proof. Assume that $v \in \mathcal{C}$ is non-negative and solves $\mathcal{P}(\pi)$. This implies that $v(0) = \psi_2\big(q(0)\big) - \tau(0) \geq 0$. Since, from Assumption 1.2.2, a trader of type $\theta = 0$ has no access to a profitable outside option, he participates. From Proposition 1.3.1 it must then hold that $\tau(0) \geq C\big(q(0)\big)$, which, together with our first observation, implies that $\psi_2\big(q(0)\big) \geq C\big(q(0)\big)$. By Assumption 1.2.1, this relation can only hold for $q(0) = 0$. From $\psi_2(0) = C(0) = 0$ we infer that both $v(0) = -\tau(0) \geq 0$ and $\tau(0) \geq C(0) = 0$ must hold, implying $v(0) = 0$. $\square$

Lemma 1.3.3. *There exists $\overline{q} \geq 0$ such that if $v \in \mathcal{C}$ solves $\mathcal{P}(\pi)$, then $|\partial v| \leq \overline{q}$.*

Proof. From Assumption 1.2.3 and the compactness of Θ we have that the mapping $q \mapsto i(\theta, v, q)$ tends to $-\infty$ as $|q| \to \infty$ uniformly on Θ for $v \geq 0$. From Proposition 1.3.1 $i\big(\theta, v(\theta), v'(\theta)\big)$ must be non-negative for all participating types, which concludes the proof. $\square$

As $\overline{q}$ could depend on π, we define

$$\mathcal{A}(\pi) := \{v \in \mathcal{C} \mid\ v \geq 0, v(0) = 0, |\partial v| \leq \overline{q}\}$$

as new admissibility set for problem $\mathcal{P}(\pi)$. The previous results show that if we replace $\mathcal{C}$ by $\mathcal{A}(\pi)$ in the definition of $\mathcal{P}(\pi)$, the solution to the problem does not change.

Corollary 1.3.4. *The admissible set $\mathcal{A}(\pi) \subset \mathcal{C}$ of Problem $\mathcal{P}(\pi)$ is uniformly bounded and uniformly equicontinuous.*

Proof. From Lemmas 1.3.2 and 1.3.3 we have that the quantity $\max_{\theta\in\Theta}\big\{u_0(\theta;\pi)\big\} + \overline{q}\|\Theta\|$ is an upper bound for any admissible choice of v. Furthermore, Lemma 1.3.3 guarantees that for any $v \in \mathcal{A}(\pi)$ it holds that $|\partial v| \leq \overline{q}$. In other words, $\mathcal{A}(\pi)$ is composed of convex functions whose subdifferentials are uniformly bounded, hence $\mathcal{A}(\pi)$ is uniformly equicontinuous. $\square$

Notice that, when it comes to determining quantities and prices for trader types who do participate, Proposition 1.3.1 results in the dealer having to solve the problem

$$\widetilde{\mathcal{P}}(\pi) := \begin{cases} \sup_{v\in\mathcal{A}(\pi)} \int_\Theta \big(i\big(\theta, v(\theta), v'(\theta)\big)\big)_+ f(\theta)\mathrm{d}\theta \\ \text{s.t. } v(\theta) \geq u_0(\theta;\pi) \text{ for all } \theta \in \Theta. \end{cases}$$

The last auxiliary result that we need is the following proposition, whose proof is a direct consequence of Fatou's Lemma, together with Lemmas 1.3.2 and 1.3.3.

Proposition 1.3.5. *The mapping $v \mapsto \int_\Theta \big(i(\theta, v(\theta), v'(\theta))\big)_+ f(\theta)\mathrm{d}\theta$ is upper semi-continuous in $\mathcal{A}(\pi)$ with respect to uniform convergence.*

We are now ready to prove our first main result:

Proof of Theorem 1.2.5. Assume that $\mathcal{A}(\pi) \cap \{ v \in \mathcal{C} \mid v(\cdot) \geq u_0(\cdot;\pi)\}$ is non-empty and consider a maximizing sequence $\{\widetilde{v}_n\}_{n\in\mathbb{N}}$ of Problem $\widetilde{\mathcal{P}}(\pi)$. From Corollary 1.3.4 we have that, passing to a subsequence if necessary, there exists $\widetilde{v} \in \mathcal{A}(\pi)$ such that $\widetilde{v}_n \to \widetilde{v}$ uniformly for $n \to \infty$. A direct application of Proposition 1.3.5 yields that $\widetilde{v}$ is a solution to $\widetilde{\mathcal{P}}(\pi)$. To finalize the proof we must construct from $\widetilde{v}$ a solution to Problem $\mathcal{P}(\pi)$. To this end, let us define the sets

$$\Theta_- := \left\{ \theta \in \Theta \,\middle|\, i\big(\theta, \widetilde{v}(\theta), \widetilde{v}'(\theta)\big) < 0 \right\} \quad \text{and} \quad \Theta_+ := \Theta_-^c.$$

It is well known that if a sequence of convex functions converges uniformly (to a convex function), then there is also uniform convergence of the derivatives wherever they exist, which is almost everywhere. This fact, together with the continuity of the mappings $\theta \mapsto \widetilde{v}(\theta)$ and $(\theta, v, q) \mapsto i(\theta, v, q)$, implies that Θ_- is the union of disjoint open intervals:

$$\Theta_- = \bigcup_{j=1}^{\infty} (a_j, b_j).$$

Define, for each $j \in \mathbb{N}$,

$$\widetilde{v}_{a,j} := \inf \{ q \mid q \in \partial\widetilde{v}(a_j)\} \quad \text{and} \quad \widetilde{v}_{b,j} := \sup \{ q \mid q \in \partial\widetilde{v}(b_j)\}$$

and consider the support lines to the graph of $\widetilde{v}$ at a_j and b_j given by

$$l_j(\theta) = \widetilde{v}(a_j) + \widetilde{v}_{a,j}(\theta - a_j) \quad \text{and} \quad L_j(\theta) = \widetilde{v}(b_j) + \widetilde{v}_{b,j}(\theta - b_j),$$

respectively. Let $c_j \in (a_j, b_j)$ be, for each $j \in \mathbb{N}$, the unique solution to the equation $l_j(\theta) = L_j(\theta)$ and define on $(a_j, b_j) =: \Theta_j$

$$v_j^*(\theta) := \begin{cases} l_j(\theta), & a_j < \theta \leq c_j; \\ L_j(\theta), & c_j < \theta < b_j. \end{cases}$$

Finally define

$$v^*(\theta) := \begin{cases} \widetilde{v}(\theta), & \theta \in \Theta_+; \\ v_j^*(\theta), & \theta \in \Theta_j,\ j \in \mathbb{N}, \end{cases}$$

then v^* is a solution to Problem $\mathcal{P}(\pi)$ and $\Theta_e(v^*) = \Theta_-$, which concludes the proof. $\square$

Remark 1.3.6. *If the dealer can profitably match all agents' outside options, then the quasi-uniqueness of a solution to Problem $\mathcal{P}(\pi)$ is in fact uniqueness. Indeed, in such a case*

$$\big(i(\theta, v(\theta), v'(\theta))\big)_+ = \big(i(\theta, v(\theta), v'(\theta))\big)$$

and problem $\widetilde{\mathcal{P}}(\pi)$ is, by Assumption 1.2.3, one of maximizing a strictly concave, coercive functional over a convex set that is closed w.r.t. uniform convergence. In the general case, we construct the quasi-unique solution in Section 1.4.2. Assumption 1.2.3 remains crucial, since it guarantees that the maximization problems through which we define the optimal quantities have unique maximizers.

1.4 The impact of a crossing network

In this section we look at the impact that a CN has on the spread, on participation and on the traders' welfare. In order to do so, we provide a characterization of the solution to Problem $\mathcal{P}(\pi)$. It should be noted that, given the restriction of candidate solutions to $\mathcal{C}$, we cannot simply make use of the Euler-Lagrange equations to solve the variational problem, since the said equations are only satisfied when the constraints do not bind.

1.4.1 A benchmark without a crossing network

We first analyze the benchmark case where the traders do not have access to a CN. The corresponding dealer's problem is denoted by $\mathcal{P}_o$. Recall that all trader types have a zero reservation utility, which the dealer is able to match costlessly by offering the contract $(0,0)$. The point of making this normalization is to simplify the constraints in the dealer's optimization problem. This will not be possible in the presence of a CN since, even if the dealer were able to match the utility that investors enjoy if they trade in the CN, this would in general not be costless.

We take a Lagrange-multiplier approach to provide a characterization of the solution to Problem $\mathcal{P}_o$. To this end, let us introduce the following definition:

$$I[v] := \int_\Theta i\big(\theta, v(\theta), v'(\theta)\big) f(\theta) \mathrm{d}\theta.$$

Let $BV_+(\Theta)$ be the space of non-negative functions of bounded variation $\gamma : \Theta \to \mathbb{R}_+$, which we place in duality with $C(\Theta, \mathbb{R})$, the space of real-valued continuous functions on Θ, via the standard pairing

$$\langle v, \gamma \rangle := \int_\Theta v(\theta) \mathrm{d}\gamma(\theta)$$

for $v \in C(\Theta, \mathbb{R})$, where $d\gamma$ is the distributional derivative of γ. Furthermore, it follows from *Pontryagin's Maximum Principle* and the fact that f is a probability density function that there is no loss of generality in assuming that γ is absolutely continuous and that $\gamma(\overline{\theta}) = 1$. The Lagrangian for the dealer's problem is

$$\mathcal{L}(v, \gamma) := I[v] + \langle v, \gamma \rangle, \quad v \in \mathcal{C},$$

with corresponding Karush-Kuhn-Tucker conditions

$$\langle v, \gamma \rangle = 0 \quad \text{and} \quad \mathrm{d}\gamma(\theta) = 0 \Rightarrow v(\theta) > 0. \tag{1.4.1}$$

The next result is the formalization of the *vox populi* saying that "quality does not jump". Regularity properties of the solutions to variational problems subject to convexity constraints were studied in [CLR01], and their methodology can be directly adapted to prove the following result.

Proposition 1.4.1. *If $v \in \mathcal{C}$ is a stationary point of $\mathcal{L}(v, \gamma)$, then $v \in C^1(\Theta)$.*

The fact that, at the optimum, the mapping $\theta \mapsto v'(\theta)$ is continuous, implies that q is also a continuous function of the types. This will prove to be extremely useful, specially in the presence of a CN. If we integrate by parts, then $\mathcal{L}(v,\gamma)$ can be transformed into

$$\Sigma(q,\gamma) := \int_\Theta \left(\Big(\theta + \frac{F(\theta)-\gamma(\theta)}{f(\theta)}\Big)\psi_1\big(q(\theta)\big) - \widetilde{C}\big(q(\theta)\big)\right) f(\theta)\mathrm{d}\theta,$$

where $q(\theta) = \psi_1^{-1}\big(v'(\theta)\big)$, as in Equation (1.2.1), and $\widetilde{C}(q) := C(q) - \psi_2(q)$. The idea now is to maximize the mapping

$$q \mapsto \sigma(\theta, q, \Gamma) := \left(\theta + \frac{F(\theta)-\Gamma}{f(\theta)}\right)\psi_1(q) - \widetilde{C}(q)$$

pointwise, for a given fixed Γ (in the sequel we use Γ whenever we are dealing with an arbitrary but fixed value of γ). From Assumption 1.2.3 it follows that we can write down the unique maximizer as

$$l(\theta,\Gamma) := K^{-1}\left(\frac{F(\theta) + \theta\, f(\theta) - \Gamma}{f(\theta)}\right),$$

where $K(q) := \widetilde{C}'(q)/\psi_1'(q)$. For each $\theta \in \Theta$ and $\Gamma \in [0,1]$, the quantity $l(\theta,\Gamma)$ is a candidate for the optimal $q(\theta)$ and convexity (or incentive compatibility) is verified if the mapping $\theta \mapsto l(\theta,\Gamma)$ is increasing. The crux is to determine the Lagrange multiplier γ. In the sequel we denote $\Theta_o := \Theta_0(v_o^*)$, where v_o^* solves Problem $\mathcal{P}_o$. In other words, if $\theta \in \Theta_o$, then $q(\theta) = T(\theta) = v(\theta) = 0$.

From Lemma 1.3.2 we have that, unless $v(\underline{\theta}) = 0$, the quantity $q(\underline{\theta}) < 0$ and the complementary–slackness condition imply that $\gamma(\theta) = 0$ for $\theta \in [\underline{\theta}, \widetilde{\theta})$ for some $\widetilde{\theta} > \underline{\theta}$. The left endpoint $\underline{\theta}_0$ of Θ_o is then determined by solving the equation

$$K^{-1}\left(\theta + \frac{F(\theta)}{f(\theta)}\right) = 0.$$

Furthermore, since v must be convex, once $v(\hat{\theta}) > 0$, then $v(\theta) > 0$ for all $\theta > \hat{\theta}$. This implies that the right endpoint $\overline{\theta}_0$ of Θ_o is determined by solving the equation

$$K^{-1}\left(\theta - \frac{1-F(\theta)}{f(\theta)}\right) = 0.$$

The quantities $F(\theta)/f(\theta)$ and $(1-F(\theta))/f(\theta)$ are know as the *hazard rates*, and sufficient conditions for the mapping $\theta \mapsto l(\theta,\Gamma)$ to be non-decreasing are

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{F(\theta)}{f(\theta)}\right) \geq 0 \geq \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{1-F(\theta)}{f(\theta)}\right),$$

see e.g. [BMR00] for a discussion on this condition.

Let us assume that we have determined Θ_o. What remains is to connect the participation constraint with the spread. Differentiating Equation (1.2.2) and noting that $v'(\theta) = \psi_1(q(\theta))$, we have that

$$\tau'(\theta) = q'(\theta)\big(\theta\psi_1'(q(\theta)) + \psi_2'(q(\theta))\big).$$

Observe that $\tau'(\underline{\theta}_0)$ and $\tau'(\overline{\theta}_0)$ are in fact $T'(0_-)$ and $T'(0_+)$, since by construction $q(\underline{\theta}_0) = q(\overline{\theta}_0) = 0$. If we define $\phi_1 := \psi_1'(0)$ and $\phi_2 := \psi_2'(0)$, then we have that the spread is given by the expressions

$$t(0_-) = q'(\underline{\theta}_{0-})\big(\underline{\theta}_0\phi_1 + \phi_2\big) \quad \text{and} \quad t(0_+) = q'(\overline{\theta}_{0+})\big(\overline{\theta}_0\phi_1 + \phi_2\big). \tag{1.4.2}$$

Our objective in Section 1.4.2 is to compare the values above to those obtained in the presence of a CN.

Before we proceed, we present two examples so as to illustrate the use of the methodology described hitherto. The first revisits [MR78]. The second is slightly more advanced. We shall use it below to illustrate the complex structure of equilibrium pricing schedules and utilities in the presence of CNs.

Example 1.4.2. *Let us assume that $\Theta = [-r, r]$ for some $r > 0$, that types are uniformly distributed and that*

$$u(\theta, q) = \theta q.$$

We also set $C(q) = 0.5\, q^2$. By direct computation we find that $\underline{\theta}_0 = -\frac{r}{2}$ and $\overline{\theta}_0 = \frac{r}{2}$. Since a trader of type $\theta \in \Theta_o$ is brought down to reservation utility and hence trades $q(\theta) = 0$, the expression

$$q(\theta) = \theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)} = 2\theta + r - 2r\gamma(\theta)$$

implies that the Lagrange multiplier is

$$\gamma(\theta) = \begin{cases} 0, & \theta < \underline{\theta}_0, \\ \frac{1}{2} + \frac{\theta}{r}, & \theta \in \Theta_o, \\ 1, & \theta > \overline{\theta}_0. \end{cases}$$

In particular, $q'(\underline{\theta}_{0-}) = q'(\overline{\theta}_{0+}) = 2$ and hence $t(0_-) = -r$ and $t(0_+) = r$. Thus, the spread increases linearly in the highest/lowest type.

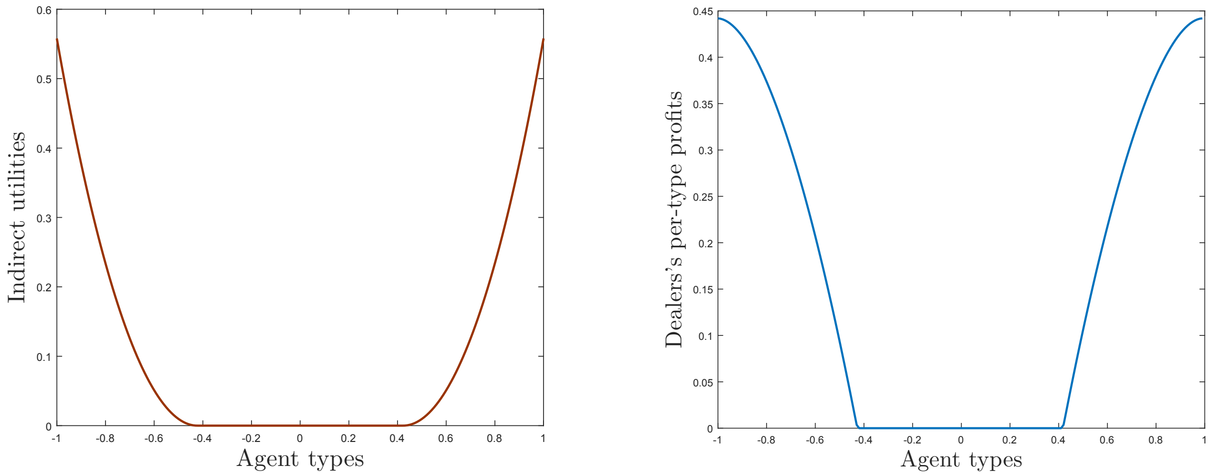
Example 1.4.3. *Let us assume that the distribution of types over $\Theta = [-1, 1]$ is given by $f(\theta) = (2\theta + 3)/4$ for $\theta \in [-1, 0)$ and $f(\theta) = (3 - 2\theta)/4$ for $\theta \in [0, 1]$; that $C(q) = 0.5\, q^2$ and that $u(\theta, q) - \tau = \theta \cdot q + 0.25\, q^2 - \tau$. It is straightforward to show that the conditions on the hazard rates are satisfied and that*

$$K^{-1}\left(\theta + \frac{F(\theta)}{f(\theta)}\right) = 2\left[\frac{3\theta^2 + 6\theta + 2}{2\theta + 3}\right] \quad \textit{and} \quad K^{-1}\left(\theta - \frac{1 - F(\theta)}{f(\theta)}\right) = 2\left[\frac{3\theta^2 - 6\theta + 2}{2\theta - 3}\right].$$

Furthermore, $\Theta_o \approx \big[-0.423, 0.423\big]$. For the spread, we have that $t(0_-) = q'(\underline{\theta}_0)\underline{\theta}_0 \approx -1.359$ and $t(0_+) = q'(\overline{\theta}_0)\overline{\theta}_0 \approx 1.359$. In order to obtain v we integrate q (since $\psi_1(q) = q$) and take into account that $v \equiv 0$ over Θ_o. We plot the graph of v_o in Figure 1.1, as well as the per-type profits of the dealer.

1.4.2 Introducing a crossing network

Let us now analyze the dealer's problem when the market participants have access to a CN that yields a trader of type θ the expected utility $u_0(\theta; \pi)$. In this setting it is no



(a) Indirect Utilities (b) Profits

Figure 1.1: An example without a CN

longer without loss of generality to assume that all traders participate in the DM, given that enforcing participation (which can be done thanks to Assumption 1.2.6) may result in losses to the dealer. The latter may, as a consequence, choose to abstain from trading with a set of types $\Theta_e(v)$ by offering an incentive-compatible book whose corresponding indirect–utility function lies strictly under $u_0(\theta;\pi)$ for $\theta \in \Theta_e(v)$. The resulting problem for the dealer would be

$$\mathcal{P}(\pi) = \sup_{v \in \mathcal{C}} \int_\Theta \Big(\theta\, v'(t) - v(t) - \widetilde{K}\big(v'(\theta)\big)\Big) \mathbb{1}_{\{\Theta_e^c(v)\}}(\theta) f(\theta) \mathrm{d}\theta.$$

Dealing with the presence of the zero–one indicator function $\mathbb{1}_{\{\Theta_e^c\}}$ is quite cumbersome (see e.g. [HMB11]), since its domain of definition may change with different book choices. In contrast to the setting studied in [HMB11], however, here the CN is passive. This lack of non-cooperative-games component allows for an alternative way to proceed. To this end, we make use of the following *accounting trick*, which was introduced in [Jul00]: Let us assume that the dealer had access to a fictitious market such that the unwinding costs from trading in it, denoted in the sequel by C_c, satisfy $C_c(q(\theta)) = \tau(\theta)$ for almost all $\theta \in \Theta$. In this way, we may again assume that the dealer trades with all market participants, but now his costs of unwinding are given by the function $\mathbb{C} : \mathbb{R} \to \mathbb{R}$ defined as

$$\mathbb{C}(q) := \min\left\{C(q), C_c(q)\right\}, \quad q \in \mathbb{R}.$$

In terms of incentives, nothing is distorted by introducing the cost function $\mathbb{C}$, but we must identify the points where there is switching from using C to using C_c and vice versa. These switching points will determine the regions of market segmentation.

If we define, for any traded quantity q, the function $\widetilde{\mathbb{C}}(q) := \mathbb{C}\big(q\big) - \psi_2\big(q\big)$, then we may re-use the machinery from Section 1.4.1 with minor modifications;[5] namely, denoting

[5] Observe that Assumptions 1.2.2 and 1.2.6 imply that $\widetilde{\mathbb{C}}$ satisfies Assumption 1.2.3.

by $\mathbb{I}$ the energy corresponding to the cost function $\mathbb{C}$, we may write the Lagrangian of the dealer's problem as

$$\mathbb{L}(v,\gamma) := \mathbb{I}[v] + \langle v - u_0(\cdot;\pi), \gamma\rangle,$$

with the corresponding complementary–slackness conditions. From here we may proceed as in Section 1.4.1 to find the quantities that the dealer will choose to offer. Strictly speaking we should find the pointwise maximizer in q of the expression

$$\left(\theta + \frac{F(\theta) - \Gamma}{f(\theta)}\right)\psi_1(q) - \mathbb{K}(q), \tag{1.4.3}$$

where $\mathbb{K}(q) := \widetilde{\mathbb{C}}(q) - \psi_2(q)$. This may fortunately be avoided, given that whenever $\mathbb{C}(q) = C_c(q)$, the participation constraint binds and $q(\theta) = q_c(\theta)$. Before proceeding to the proof of Theorem 1.2.7, we study the mechanism used by the dealer to choose between excluding types, matching the CN, and fully servicing them by offering strictly positive rents.

Whenever the participation constraint does not bind, the dealer selects the quantity to be chosen via the pointwise maximization of the mapping $q \mapsto \sigma(\theta, q, \Gamma)$. What makes the current problem trickier than the case without a CN is that now we must pay more attention to the evolution of the multiplier γ. If we compare $l(\theta,0)$ and $l(\theta,1)$ to $q_c(\theta)$, we may pinpoint the set where the participation constraint may bind. Observe that $\{l(\theta,1), \theta \in \Theta\}$ and $\{l(\theta,0), \theta \in \Theta\}$ are the sets of the lowest and highest quantities the dealer may offer in an individually-rational way. Hence, as long as $l(\theta,1) \le q_c(\theta) \le l(\theta,0)$ there is the possibility of *profitable matching*.

There might be instances where the participation constraint is binding for some type $\theta \in \Theta$, i.e. $\big(q(\theta), \tau(\theta)\big) = \big(q_c(\theta), \tau_c(\theta)\big)$, and $\tau_c(\theta) - C\big(q_c(\theta)\big) < 0$. In such cases $\mathbb{C}\big(q_c(\theta)\big) = C_c\big(q_c(\theta)\big)$ and $\theta \in \Theta_e(v)$ for the corresponding indirect–utility function, and we say there is *exclusion*.

Remark 1.4.4. *It is at this point that the quasi-uniqueness mentioned in Remark 1.3.6 can be addressed. The principal's problem $\mathcal{P}_{\mathbb{C}}(\pi)$ using the cost function $\mathbb{C}$ results in the condition*

$$\big(i(\theta, v(\theta), v'(\theta))\big)_+ = \big(i(\theta, v(\theta), v'(\theta))\big)$$

being trivially satisfied. As a consequence, problem $\mathcal{P}_{\mathbb{C}}(\pi)$ admits a unique solution. The latter coincides, by construction, with the solution to $\mathcal{P}(\pi)$ whenever $\mathbb{C}(q(\theta)) = C(q(\theta))$. The caveat is that the solution to problem $\mathcal{P}_{\mathbb{C}}(\pi)$ is blind towards what is offered to excluded types, since here their outside option is costlessly matched (they are effectively reserved). Constructing incentive-compatible contracts for the excluded types is, thanks to the convexity of the indirect–utility function, relatively simple. For instance if an interval of types (θ_1, θ_2) were excluded (but θ_1 and θ_2 participated) one could consider any two supporting lines to the graph of $v(\cdot;\pi)$ at $(\theta_1, v(\theta_1;\pi))$ and $(\theta_2, v(\theta_2;\pi))$. From the resulting indirect–utility function on (θ_1, θ_2) one could extract the corresponding quantities and prices. The resulting global convexity of the indirect–utility function offered by the principal would imply that all incentives would remain unchanged. Whether the principal would suffer losses from the contracts offered to types on (θ_1, θ_2) would be irrelevant, since the corresponding agents do not participate.

As mentioned above, here it is not necessary to determine $\gamma(\theta)$ in order to determine $q(\theta)$. On the other hand, however, if we interpret γ as the shadow cost of satisfying the participation constraint, we may wish to identify the multiplier so as to have a measure of the impact of the CN on the dealer's profits. The following result, which deals with points where there is switching between matching and fully servicing, extends Proposition 1.4.1.

Proposition 1.4.5. *For $\pi \in \mathbb{R}^2$ given, let $\widetilde{\theta} \in \Theta$ be such that there exists $\epsilon > 0$ such that $v(\theta;\pi) = u_0(\theta;\pi)$ on $(\widetilde{\theta}-\epsilon, \widetilde{\theta}]$ and $v(\theta;\pi) > u_0(\theta;\pi)$ on $(\widetilde{\theta}, \widetilde{\theta}+\epsilon]$. Furthermore, assume that*

$$\int_{\widetilde{\theta}-\epsilon}^{\widetilde{\theta}} \big(\tau(\theta) - C(q(\theta))\big) f(\theta)\mathrm{d}\theta > 0,$$

where $\{\big(q(\theta),\tau(\theta)\big), \theta \in \Theta\}$ implements $v(\cdot;\pi)$. In other words, there is profitable matching on $(\widetilde{\theta}-\epsilon, \widetilde{\theta}]$ and the dealer fully services types on $(\widetilde{\theta}, \widetilde{\theta}+\epsilon]$. Then $\partial v(\widetilde{\theta};\pi)$ is a singleton. The result also holds if the order of the matching and full-servicing intervals is switched.

The rationale behind Proposition 1.4.5 is that, as long as the dealer is able to match the traders' outside option without incurring a loss, it is possible to normalize the latter to zero and directly apply Proposition 1.4.1. This is, naturally, not the case when matching u_0 results in losses. We put Proposition 1.4.5 to work in Example 1.4.9.

Before moving on, we present below a modification to Example 1.4.3 that shows how even agents without access to a non-trivial outside option benefit from the presence of the CN and that the optimal Lagrange multiplier need not be continuous.

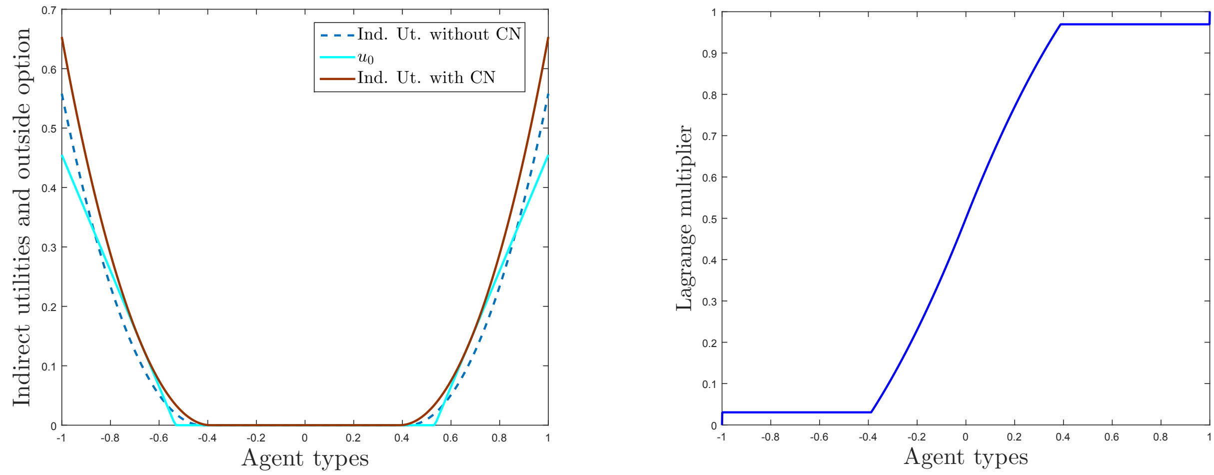
Example 1.4.6. *Let f, Θ, C and u be as in Example 1.4.3 and assume that the CN offers the traders the following expected profits:*

$$u_0(\theta;(3.2,3.2)) = \begin{cases} -0.975\theta - 0.52, & \text{if } \theta \leq -\frac{8}{15}; \\ 0.975\theta - 0.52, & \text{if } \theta \geq \frac{8}{15}; \\ 0, & \text{if } \theta \in (-\frac{8}{15}, \frac{8}{15}). \end{cases}$$

Matching this outside option would require the dealer to offer the contracts $(\pm 0.975, 0.52)$. This is profitable, hence the indirect utility never lies below u_0. To illustrate this, we have plotted the indirect–utility function in Figure 1.2(a). It strictly dominates the one plotted in Figure 1.1(a) for all types who earn positive profits. The smooth pasting condition ($l(\theta,\gamma(\theta)) = q_c(\theta)$ where v touches u_0, i.e. in ± 0.675) determines the optimal Lagrange multiplier, namely $\gamma(-1) = 0$ and $\gamma \equiv 0.030$ on $(-1,-0.389]$. For positive types we obtain symmetrically $\gamma(1) = 1$ and $\gamma \equiv 0.970$ on $[0.389, 1)$. The new spread, given by $\big(t(0_-), t(0_+)\big) = (-1.282, 1.282)$, is strictly smaller than in the case without a CN.

The following result will prove to be essential for the results in Section 1.5. It guarantees, by virtue of Assumption 1.2.2, that our notion of the spread is well defined in the presence of a CN, and could be loosely summarized by saying that the first (in terms of moving away from $\theta = 0$) types to earn positive utility trade in the DM.

Lemma 1.4.7. *For any $\pi \in \mathbb{R}^2$, there exists $\epsilon = \epsilon(\pi) > 0$ such that the types that belong to $(\underline{\theta}_0(\pi)-\epsilon, \underline{\theta}_0(\pi)) \cup (\overline{\theta}_0(\pi), \overline{\theta}_0(\pi)+\epsilon)$ are fully serviced.*



(a) Indirect Utilities (b) Lagrange Multiplier

Figure 1.2: An example without exclusion

Proof. Let us denote by $\hat{\theta}$ the largest positive solution to the equation $u_0(\theta;\pi) = 0$. If there exists $\eta > 0$ such that types on $(\hat{\theta}, \hat{\theta}+\eta)$ can be matched profitably, then the result follows either because $\overline{\theta}_0(\pi) < \hat{\theta}$ or because $\overline{\theta}_0(\pi) = \hat{\theta}$ and the types on $(\hat{\theta}, \hat{\theta}+\epsilon)$, for some $0 < \epsilon \leq \eta$, are fully serviced. Let us now assume that such an η does not exist, we claim then that $\overline{\theta}_0(\pi) < \hat{\theta}$ must hold. Proceeding by the way of contradiction, let us assume that $\overline{\theta}_0(\pi) = \hat{\theta}$ (which is equivalent to $\overline{\theta}_0(\pi) \geq \hat{\theta}$) and that there exists $\delta > 0$ such that $(\hat{\theta}, \hat{\theta}+\delta) \subset \Theta_e(\pi)$. This configuration can be improved upon as follows: let $a > 0$ be such that $\hat{\theta}-a > 0$. By construction $l(\hat{\theta}-a, \gamma(\hat{\theta}-a)) = 0$. Let us fix $\gamma(\theta) \equiv \gamma(\hat{\theta}-a) =: \Gamma(a)$ for $\theta \in (\hat{\theta}-a, \theta_a)$, where θ_a is the solution to $v_a(\theta) = u_0(\theta;\pi)$ on $(\hat{\theta}-a, \overline{\theta}]$ if it exists or $\theta_a = \overline{\theta}$ otherwise, given that we denote by v_a the indirect–utility function corresponding to $\Gamma(a)$. In particular, $\theta_a > \hat{\theta}$ and $l(\theta, \Gamma(a)) > 0$ for $\theta \in (\hat{\theta}-a, \theta_a)$.

We now have that types $\theta \in (\hat{\theta}-a, \theta_a)$ are fully serviced. By Assumption 1.2.2, $v_a'(\hat{\theta}-a) = 0 < u_0'(\hat{\theta};\pi)$; therefore, there exists $a_1 > 0$ such that for all $a \leq a_1$ it holds that $\theta_a < \hat{\theta}+\delta$. If we could show that there exists $a \leq a_1$ such that the principal could offer types in $(\hat{\theta}-a, \theta_a)$ the quantities $q_a(\theta) = l(\theta, \Gamma(a))$ at a profit, we would contradict the optimality of $\overline{\theta}_0(\pi)$ and the proof would be complete, since incentives above θ_a would not be distorted and the principal's profits would strictly increase. In order to do so, observe that the principal's typewise profit when offering $q_a(\theta)$ is

$$P(\theta) := \theta\psi_1(q_a(\theta)) + \psi_2(q_a(\theta)) - v_a(\theta) - C\big(q_a(\theta)\big).$$

In particular, $P(\hat{\theta}-a) = 0$ and

$$\begin{aligned}
&P'(\hat{\theta}-a) \\
&= \psi_1(q_a(\hat{\theta}-a)) + (\hat{\theta}-a)\psi_1'(q_a(\hat{\theta}-a))q_a'(\hat{\theta}-a) + v_a'(\hat{\theta}-a) - \widetilde{C}'\big(q_a(\hat{\theta}-a)\big)q_a'(\hat{\theta}-a) \\
&= \psi_1(0) + (\hat{\theta}-a)\psi_1'(0)q_a'(\hat{\theta}-a) + v_a'(\hat{\theta}-a) - \widetilde{C}'\big(0\big)q_a'(\hat{\theta}-a) \\
&= (\hat{\theta}-a)\psi_1'(0)q_a'(\hat{\theta}-a).
\end{aligned}$$

The third equality follows because by construction $v_a'(\hat{\theta}-a) = 0$; by assumption $\psi_1(0) = 0$ and, from Assumption 1.2.3, $\widetilde{C}'\big(0\big) = 0$. Furthermore, since ψ_1 is strictly increasing and $q_a'(\hat\theta - a) > 0$, then $P'(\hat\theta - a) > 0$. Therefore, there exists $b > 0$ such that $P(\theta) > 0$ if $\theta \in (\hat\theta - a, \hat\theta - a + b)$. As a consequence, if $a < a_1$ is small enough, then $P(\theta) > 0$ for $\theta \in (\hat\theta - a, \theta_a)$, as required. □

We are now in the position to present the proof of our second main result.

Proof of Theorem 1.2.7. (1) Observe that if π is such that $\big(\underline{\theta}_0(\pi), \overline{\theta}_0(\pi)\big) = \Theta_0(\pi) \subset \Theta_o$, then the result follows immediately from Lemma 1.4.7. If we revert the inclusion, two situations are possible, since the addition of the CN-constraint to Problem $\mathcal{P}_o$ may or may not bind for some types. The latter case being trivial, let us look at the case where there is a point $\theta_a > \overline{\theta}_0$ on which it holds that $v_o(\theta_a) = u_0(\theta_a; \pi)$ and such that $v_o(\theta) > u_0(\theta; \pi)$ for $\theta < \theta_a$ and vice versa for $\theta > \theta_a$. The Lagrange multiplier γ_m is active on $(\theta_a, \overline{\theta}]$, which implies that $\gamma_m(\theta_a) < 1$. We know from [Jul00, p. 9] that for all θ such that $l(\theta, \Gamma) > 0$, the latter is decreasing in Γ. As a consequence, the root of the equation

$$K^{-1}\left(\theta + \frac{F(\theta) - \gamma_m(\theta_a)}{f(\theta)}\right) = 0$$

is strictly smaller than that of $l(\theta, 1) = 0$, which yields the desired result.

(2) Let us denote by $t_o(0_-)$ and $t_o(0_+)$ the best bid and ask prices without the presence of a CN and by $t_m(0_-)$ and $t_m(0_+)$ the corresponding marginal prices with one; thus,

$$t_o(0_-) = q_o'(\underline{\theta}_{0,o-})\big(\underline{\theta}_{0,o}\phi_1 + \phi_2\big) \text{ and } t_o(0_+) = q_o'(\overline{\theta}_{0,o+})\big(\overline{\theta}_{0,o}\phi_1 + \phi_2\big)$$

and

$$t_m(0_-) = q_m'(\underline{\theta}_{0,m-})\big(\underline{\theta}_{0,m}\phi_1 + \phi_2\big) \text{ and } t_m(0_+) = q_m'(\overline{\theta}_{0,m+})\big(\overline{\theta}_{0,m}\phi_1 + \phi_2\big).$$

From Part (1) we know that $\underline{\theta}_{0,o} \le \underline{\theta}_{0,m}$ (both negative) and $\overline{\theta}_{0,m} \le \overline{\theta}_{0,o}$ (both positive) and, since ϕ_1 and ϕ_2 do not depend on the presence of the CN, all we have left to do is show that

$$q_m'(\underline{\theta}_{0,m-}) \le q_o'(\underline{\theta}_{0,o-}) \text{ and } q_m'(\overline{\theta}_{0,m+}) \le q_o'(\overline{\theta}_{0,o+}).$$

Using the well-known relation $(f^{-1})'(a) = 1/f'(f^{-1}(a))$ for the derivative of a function's inverse we have that

$$\begin{aligned}
q_m'(\underline{\theta}_{0,m-}) &= \frac{1}{K'\Big(K^{-1}\Big(\underline{\theta}_{0,m} - \frac{\gamma(\underline{\theta}_{0,m-}) - F(\underline{\theta}_{0,m})}{f(\underline{\theta}_{0,m})}\Big)\Big)} \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\theta - \frac{\gamma(\theta) - F(\theta)}{f(\theta)}\right)\Big|_{\theta = \underline{\theta}_{0,m-}} \\
&= \frac{1}{K'\big(q_m(\underline{\theta}_{0,m})\big)} \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\theta - \frac{\gamma(\theta) - F(\theta)}{f(\theta)}\right)\Big|_{\theta = \underline{\theta}_{0,m-}} \\
&= \frac{1}{K'(0)}\left(1 - \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\gamma(\underline{\theta}_{0,m-}) - F(\theta)}{f(\theta)}\right)\right)\Big|_{\theta = \underline{\theta}_{0,m}},
\end{aligned}$$

where we have used the fact that γ is constant on $(\underline{\theta}_{0,m} - \delta, \underline{\theta}_{0,m})$ for some $\delta > 0$. We may proceed analogously for the other three quantities. We have to show that

$$\begin{aligned}
&\frac{1}{K'(0)}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\gamma(\underline{\theta}_{0,m-}) - F(\theta)}{f(\theta)}\right)\Big|_{\theta = \underline{\theta}_{0,m}} \ge \frac{1}{K'(0)}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{-F(\theta)}{f(\theta)}\right)\Big|_{\theta = \underline{\theta}_{0,o}} \text{ and} \\
&\frac{1}{K'(0)}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\gamma(\overline{\theta}_{0,m+}) - F(\theta)}{f(\theta)}\right)\Big|_{\theta = \overline{\theta}_{0,m}} \ge \frac{1}{K'(0)}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{1 - F(\theta)}{f(\theta)}\right)\Big|_{\theta = \overline{\theta}_{0,o}},
\end{aligned} \tag{1.4.4}$$

which hold with equality under the assumption that $f = (\overline{\theta} - \underline{\theta})^{-1}\mathbb{1}_{\Theta}$.

(3) It follows from Part (1) that, if θ participates in the presence of the CN, then $q_o(\theta) \leq q_m(\theta)$. Assume now that the inequality $v_o(\theta) > v(\theta; \pi)$ holds for all θ in a non-empty interval (θ_1, θ_2) and $v_o(\theta_1) = v(\theta_1; \pi)$ and $v_o(\theta_2) = v(\theta_2; \pi)$. By the convexity of v_o and $v(\cdot; \pi)$, this would imply the existence of $\theta_3 \in (\theta_1, \theta_2)$ such that $v_o'(\theta) > v'(\theta; \pi)$ holds almost surely in (θ_1, θ_3). However $v_o'(\theta) = \psi_1(q_o(\theta))$, $v'(\theta; \pi) = \psi_1(q_m(\theta))$ and ψ_1 is strictly increasing; hence, this would imply that $q_o(\theta) > q_m(\theta)$ for almost all $\theta \in (\theta_1, \theta_3)$, which is a contradiction. $\square$

We finalize this section with two examples that showcase the results obtained thus far. Example 1.4.8 shows that, in the simple case where the outside option is such that the dealer will (only) exclude all high-enough (in absolute value) types, the results of Theorem 1.2.7 follow trivially.

Example 1.4.8. *Let us revisit Example 1.4.2 with an extremely steep outside option that will warrant exclusion, namely, for $r_0 < r$ let*

$$u_0(\theta) = \begin{cases} \infty, & \textit{if } \theta \in [-r, -r_0) \cup (r_0, r]; \\ 0, & \textit{otherwise}. \end{cases}$$

Recall that, for a given value Γ of the Lagrange multiplier, the corresponding quantity is

$$q(\theta; \Gamma) := 2\theta + r - 2r\Gamma.$$

In Example 1.4.2 the participation constraint does not bind for high types. In particular, $\gamma \equiv 0$ on $[-r, \underline{\theta}_0)$ and to find the left endpoint of the reserved set we set $\Gamma = 0$ and solve $2\theta + r = 0$. In the current setting, the participation constraint must bind for $\theta < -r_0$ and the multiplier will be constant on $(-r_0, \underline{\theta}_0(\Gamma))$, where

$$\underline{\theta}_0(\Gamma) := -\frac{r}{2}\big[1 - 2\Gamma\big].$$

By construction, the choice of Γ will bear no weight on the trader types that will be serviced to the left of $\theta = -r_0$, but only on how many additional low types benefit from the presence of the outside option. By integrating $q(\theta; \Gamma)$ and noting that the corresponding indirect–utility function $v(\cdot; \Gamma)$ must satisfy $v(\underline{\theta}_0(\Gamma); \Gamma) = 0$, we have

$$v(\theta; \Gamma) = \theta^2 + \theta r\big[1 - 2\Gamma\big] + \frac{r^2}{4}\big[1 - 2\Gamma\big]^2$$

for $\theta \in [-r_0, \underline{\theta}_0(\Gamma)]$. Since the indirect–utility function also satisfies $v(\theta; \Gamma) = \theta q(\theta; \Gamma) - \tau(\theta; \Gamma)$, we have that the dealer market on $[-r_0, \underline{\theta}_0(\Gamma)]$ is described by the quantity–price pairs $\big(q(\theta; \Gamma), \theta^2 - \frac{r^2}{4}\big[1 - 2\Gamma\big]^2\big)$. As a consequence, the per-type profit is

$$\Pi(\theta; \Gamma) := -\theta^2 - \frac{3}{4}r^2\big[1 - 2\Gamma\big]^2 - 2\theta r\big[1 - 2\Gamma\big],$$

where the third term on the right-hand side is positive and dominates the first two. Finally, we have that each choice of Γ will result in the dealer obtaining the aggregate profits from negative types

$$P(\Gamma) := \frac{1}{2r}\int_{-r_0}^{\underline{\theta}_0(\Gamma)} \Pi(\theta; \Gamma)\mathrm{d}\theta.$$

The mapping $\Gamma \mapsto P(\Gamma)$ is strictly concave and the first–order conditions yield that it is maximized at $\Gamma = (r - r_0)/(2r)$. As a result $\underline{\theta}_0(\Gamma) = -r_0/2$ and $v(\theta;\Gamma) = \theta^2 + r_0\theta + r_0^2/4$, which correspond to the boundary of the reserved set and the indirect–utility function for negative trader types in the problem without a CN on $[-r_0, r_0]$.

Example 1.4.9. *We stay with the basic setup of Examples 1.4.3 and 1.4.6, but now assume that $u_0(\theta;\pi) = \Big(\frac{1-\pi_+}{3}\theta^{6/5} - 0.001\Big)_+$ for $\theta \geq 0$ and $u_0(\theta;\pi) \equiv 0$ otherwise. For any type θ such that $u_0(\theta) > 0$ it holds that*

$$\big(q_c(\theta), \tau_c(\theta)\big) = \Big(\frac{2}{5}(1-\pi_+)\theta^{1/5}, \frac{2}{5}(1-\pi_+)\theta^{6/5} + \frac{1}{25}(1-\pi_+)^2\theta^{2/5} - \big(\frac{1}{3}(1-\pi_+)\theta^{6/5} - 0.001\big)_+\Big).$$

We assume $\pi = (0, 1/2)$. The first thing to notice is that the dealer's per-type profit for offering $(q_c(\theta), \tau_c(\theta))$, i.e. $\tau_c(\theta) - C(q_c(\theta)) = \theta^{6/5}/30 - \theta^{2/5}/100 + 0.001$, is negative for types $\theta \in (0.0035, 0.1667)$. On the other hand, the inequality $u_0(\theta; 1/2) \geq 0$ only holds for $\theta \geq 0.014$. Combining both arguments we see that $\Theta_e(\pi) \subset (0.014, 0.1667)$. Next we observe that the inequality

$$l(\theta, 1) = K^{-1}\Big(\theta - \frac{1 - F(\theta)}{f(\theta)}\Big) \geq \frac{\sqrt[5]{\theta}}{5}$$

holds for all $\theta \in [0.4761, 1]$.
Hence profitable matching may occur on the interval $(0.1667, 0.4761)$, over which $q(\theta) = q_c(\theta)$ and $\mathbb{C}\big(q(\theta)\big) = C\big(q(\theta)\big)$. Furthermore, Proposition 1.4.5 implies that the corresponding indirect–utility function will be differentiable at $\theta = 0.4761$. In order to obtain $v(\theta;\pi)$ for $\theta \in [0.4761, 1]$, we integrate $l(\cdot, 1)$ and determine the corresponding integration constant c by equating

$$2\int_0^{0.4761} \Big(\frac{3\theta^2 - 6\theta + 2}{2\theta - 3}\Big)\,\mathrm{d}\theta + c = \frac{1}{6}(0.4761)^{6/5} - 0.001.$$

We know from the example without a CN that $\gamma(t) = 0$ for $\theta \in [-1, -0.423)$. On $[-0.423, 0)$ the multiplier must satisfy

$$K^{-1}\Big(\theta - \frac{\gamma(\theta) - F(\theta)}{f(\theta)}\Big) = 0,$$

which results in $\gamma(\theta) = (3\theta^2 + 6\theta + 2)/4$ on the said interval. What remains to be determined is $\overline{\theta}_0$ and $\gamma(\overline{\theta}_0)$. To this end, we define the family of functions $v(\cdot;\Gamma)$ such that $v'(\theta;\Gamma) = l(\theta,\Gamma)$ whenever this quantity is positive and $v(\theta;\Gamma) = 0$ for $\theta \in [0, \theta(\Gamma)]$, where $\theta(\Gamma)$ is the solution to the equation $l(\theta,\Gamma) = 0$. Since $\gamma(0) = 0.5$, we have that $\Gamma > 0.5$.[6] In fact, $\Gamma = \gamma(\overline{\theta}_0) = 0.5105$, $\overline{\theta}_0 = 0.007$ and the intersection of $v(\cdot;\Gamma)$ and $u_0(\cdot;1/2)$ occurs at $\theta = 0.0159$.
Summarizing, the types on $[-1, -0.423) \cup (0.007, 0.0159] \cup (0.1667, 1]$ are fully serviced, those on $[-0.423, 0.007]$ are reserved and the ones that lie on $(0.0159, 0.1667)$ are excluded. The left-hand side of the spread is the same as in the example without a CN, whereas the right-hand side is $t(0_+) = 0.0281$. This is significantly smaller than in Example 1.4.3.

[6] Pasting when passing from servicing to excluding need not be smooth.

Determining $\gamma(\theta)$ *on* $(0, 0.007]$ *is relatively simple, as we again must solve* $l(\theta, \gamma(\theta)) = 0$, *which results in* $\gamma(\theta) = (-3\theta^2 + 6\theta + 2)/4$. *Finally, in order to determine* γ *on* $\Theta_e(\pi)$ *we must rewrite the virtual surplus using* $\mathbb{C}(q(\theta)) = \tau_c(\theta)$, *which results in* $\mathbb{C}(q) = (5^5/6)q^6 - (1/4)q^2 + 0.001$. *The pointwise maximization of the resulting virtual surplus must equal* $q_c(\theta) = \sqrt[5]{\theta}/5$. *After some lengthy arithmetic that we choose to spare the reader from, we obtain*

$$\gamma(\theta) = F(\theta) - f(\theta)\left[5^5 q_c(\theta)^5 - \theta\right] = F(\theta) \quad \textit{for } \theta \in \Theta_e(\pi).$$

Finally, in the profitable–matching region we solve $l(\theta, \gamma(\theta)) = \sqrt[5]{\theta}/5$ *so as to find the multiplier, which yields*

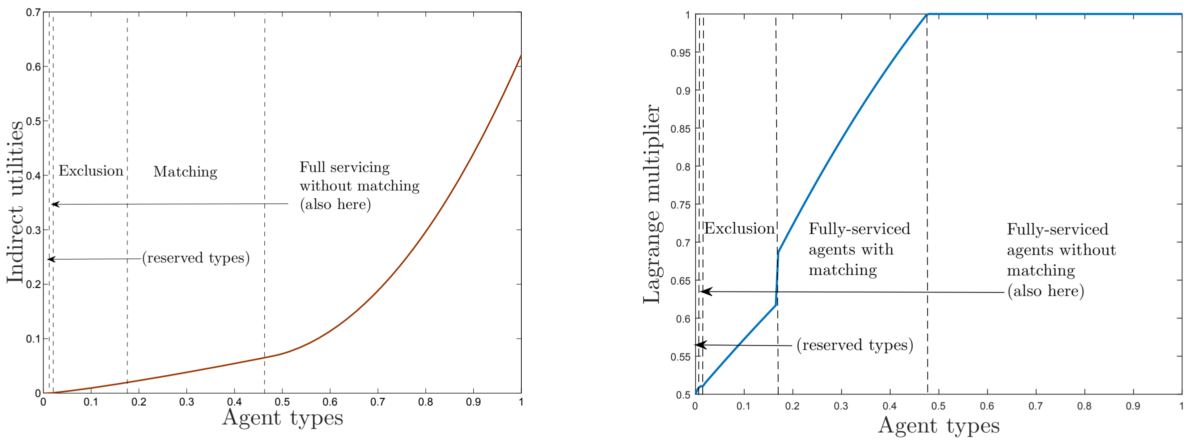
$$\gamma(\theta) = F(\theta) - f(\theta)\left[\frac{1}{10}\theta^{1/5} - \theta\right] \quad \textit{for } \theta \in [0.1667, 0.4761),$$

i.e.,

$$\gamma(\theta) = \frac{1}{10}\theta^{1/5} \cdot \frac{2\theta - 3}{4} - \frac{3\theta^2 - 6\theta - 2}{4} \quad \textit{for } \theta \in [0.1667, 0.4761).$$

Observe that, in contrast to Example 1.4.6, here $\gamma(\theta) = 1$ *for types that are strictly smaller than one. This means that the rightmost types do not profit from the introduction of the CN via changes in the quantities they are offered, but rather from changes in the corresponding prices. Intuitively speaking this has to do with how steep the outside option is for large types and, as a consequence, whether or not it will be matched over a non-trivial interval.*

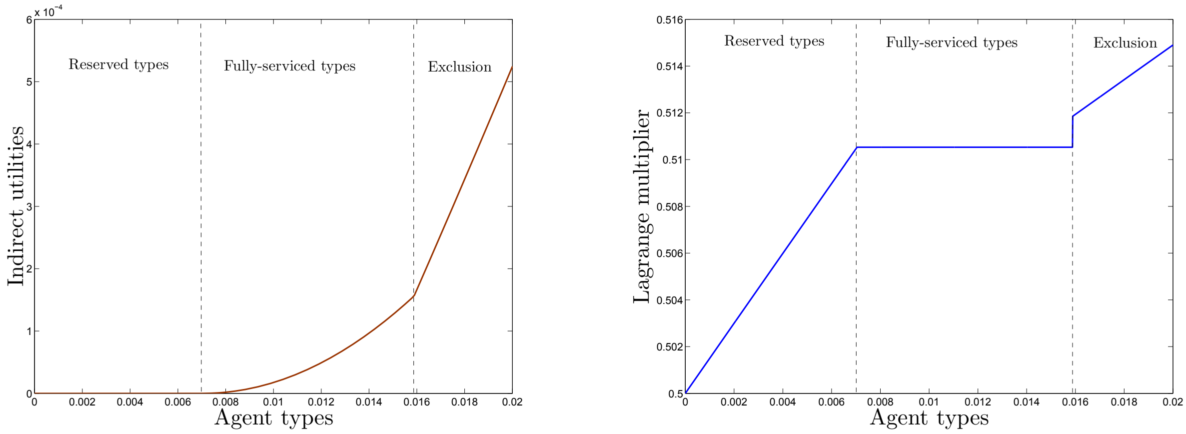
We present in Figure 1.3(a) the indirect utilities for positive types (the ones for negative types being the same as in Figure 1.1(a)). The values of γ *have been plotted in Figure 1.3(b). In Figure 1.4 we provide a magnification around small values of* θ *so as to highlight the switching between reservation, full servicing and exclusion. Observe the jump of the Lagrange multiplier at the boundary between fully-serviced and excluded types (Figure 1.4(b)) and between excluded and matched ones (Figure 1.3(b)).*



(a) Indirect Utilities (b) Lagrange Multiplier

Figure 1.3: An example with exclusion

We shall revisit this example in the upcoming section, where we look into the existence of equilibrium prices in the CN.



(a) Indirect Utilities (b) Lagrange Multiplier

Figure 1.4: An example with exclusion (magnified)

1.5 An equilibrium price in the crossing network

In this section we prove the existence of an equilibrium price π^*. We first observe that, from Assumption 1.2.8, there is no loss of generality in assuming that π^* belongs to some closed and bounded subset of $\mathbb{R}^2$, which we denote by Π. As a consequence we have that $t(0;\cdot) : \Pi \to \Pi$. The restriction of possible equilibrium prices to Π, together with Assumptions 1.2.2 and 1.2.8, yields the next result.

Lemma 1.5.1. *There exists a non-empty interval* $[\epsilon_1, \epsilon_2] \subset \Theta$ *such that*

1. $0 \in (\epsilon_1, \epsilon_2)$;
2. $u_0(\theta;\pi) = 0$ *for all* $\theta \in [\epsilon_1, \epsilon_2]$ *and all* $\pi \in \Pi$.

In the sequel we make use of the results obtained in Section 1.4.2 to show that the mapping $\pi \mapsto t(0;\pi)$ has the required monotonicity properties so as to use the following result:

Theorem 1.5.2. *(Tarski's Fixed Point Theorem,* [Tar55, Theorem 1]*) Let* $(X, \leq)$ *be a complete lattice. If* $f : X \to X$ *is order preserving, then the set of fixed points of* f *is also a (non-empty) complete lattice*[7].

We are now ready to give the proof of our third main result.

Proof of Theorem 1.2.9. Lemmas 1.4.7 and 1.5.1 guarantee that we have a well-defined spread; thus, we may decompose the analysis of the mapping $\pi \mapsto t(0;\pi)$ into that of the mappings $\pi_- \mapsto t(0_-;\pi_-)$ and $\pi_+ \mapsto t(0_+;\pi_+)$. In other words, for a given price π, the dealer's optimal response to $u_0(\cdot;\pi)$ is, modulo a normalization of γ, equivalent to the combination of his actions towards negative and positive types separately. We shall concentrate on the existence of a fixed point of the mapping $\pi_+ \mapsto t(0_+;\pi_+)$.

From Assumption 1.2.8 we have that if $\pi_{1+} < \pi_{2+}$, then $u_0(\theta;\pi_{1+}) > u_0(\theta;\pi_{2+})$ for all $\theta > 0$. If for $i = 1, 2$ it holds that $u_0(\theta;\pi_{i+}) < v_o(\theta)$ for all $\theta > 0$, then $v(\theta;\pi_{1+}) = v(\theta;\pi_{2+})$

[7]Every complete lattice is, by definition, non-empty.

on the same domain and $t(0_+;\pi_{1+}) = t(0_+;\pi_{2+})$. Next assume that $u_0(\theta;\pi_{i+}) \geq v_o(\theta)$ on a subset Θ_i of $(0,\overline{\theta}]$, for $i = 1,2$. Given that $u_0(\theta;\pi_{1+}) > u_0(\theta;\pi_{2+})$ for all $\theta > 0$, then $\overline{\theta}(\pi_1) < \overline{\theta}(\pi_2)$ and the first point $\widetilde{\theta}_1$ such that $v(\theta;\pi_{1+}) = u_0(\theta;\pi_{1+})$ holds satisfies $\widetilde{\theta}_1 < \widetilde{\theta}_2$, where the latter is the analogous to $\widetilde{\theta}_1$ in the presence of $u_0(\theta;\pi_{2+})$. The existence of $\widetilde{\theta}_1$ and $\widetilde{\theta}_2$ is guaranteed by the fact that in both cases the indirect–utility functions intersect the corresponding outside options. Arguing as in the proof of Theorem 1.2.7, Part (2), this also implies that $\overline{\theta}_0(\pi_1) < \overline{\theta}_0(\pi_2)$; hence $t(0_+;\pi_{1+}) < t(0_+;\pi_{2+})$. In other words, the mapping $\pi_+ \mapsto t(0_+;\pi_+)$ is order-preserving and, using Tarski's Fixed Point Theorem, we may conclude it has a fixed point. □

Remark 1.5.3. *The requirement of uniformly distributed types can be relaxed to the extent that if f and K are such that Conditions* (1.4.4) *are satisfied, then the required monotonicity properties still apply. Unfortunately, these conditions cannot be verified ex-ante, since they include the end points of the set of reserved traders.*

Example 1.5.4. *Let us go back to our example with exclusion, but introduce the feedback loop between the DM and the CN through the iteration $\pi_{i+1} = t(0;\pi_i)$. We initialize the recursion by setting $\pi_0 = (0,1/2)$ and $\kappa = 0.001$, which are the parameters in the aforementioned example.*

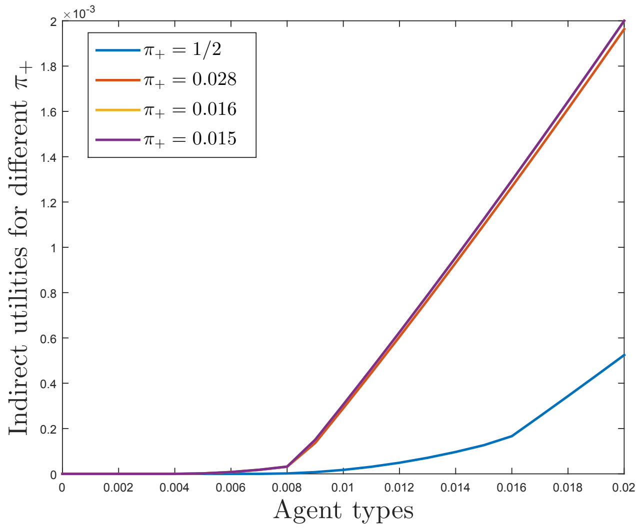


Figure 1.5: The indirect–utility functions corresponding to the iteration $\pi_{i+1} = t(0;\pi_i)$.

We observe a very swift convergence. Indeed, it takes only four iterations to reach $\|v(\cdot;\pi_i) - v(\cdot;\pi_{i+1})\|_\infty \leq 10^{-5}$ and the indirect–utility functions in the third and fourth iteration are almost indistinguishable. The equilibrium price is $\pi^* = (0,0.015)$. We present in Figure 1.5 the plots of the first four iterates. It is evident that each iteration results in a smaller set of reserved traders and in a higher indirect utility for all types. The spreads, the right endpoints of the reserved regions, the Lagrange multipliers at the right endpoint of the reserved regions and the exclusion regions are provided in Table 1.1. It is interesting to observe that, as the spread decreases to its equilibrium level, the number of trader types that are reserved decreases and the sets of excluded types grow (in terms of inclusions). This last fact obeys the fact that, when the traders have a more attractive outside option, it is harder for the dealer to match it profitably.

Table 1.1: The numbers of the feedback loop

π_+	Θ_o	Γ	$\Theta_e(\pi_+)$
$1/2$	[-0.423,0.0070]	0.5105	[0.0159, 0.1667]
0.0281	[-0.423,0.0040]	0.5061	[0.0083, 0.4872]
0.0161	[-0.423,0.0040]	0.5060	[0.0082, 0.4954]
0.0158	[-0.423,0.0040]	0.5060	[0.0082, 0.4955]

1.6 Portfolio liquidation and DP trading

In this section we present an application of our methodology to portfolio liquidation. We assume that the market participants' aim is to liquidate their current holdings on some traded asset. The sizes of the traders' portfolios are heterogeneous and saying that a trader's type is θ means that he holds θ shares of the asset prior to trading. We set $\Theta = [-1, 1]$ and $f = \frac{1}{2}\mathbb{1}_\Theta$. If a trader of type θ trades q shares for τ dollars, his utility is

$$\hat{u}(\theta, q) - \tau := -\alpha(\theta - q)^2 - \tau,$$

where $\alpha > 0$ denotes the traders' (homogeneous) sensitivity towards inventory holdings. Notice that $-\alpha\theta^2$ is the type-dependent reservation utility of a trader of type θ. If we "normalize" the said utility to zero, we may write

$$u(\theta, q) - \tau = 2\alpha\theta q - \alpha q^2 - \tau.$$

In this example the CN takes the form of a Dark Pool (DP). Choosing to trade in the latter entails two kinds of costs for the traders: On the one hand, there is a direct fixed cost $\kappa > 0$ of engaging in DP trading. On the other hand, execution in the DP is not guaranteed. We denote by $p \in [0, 1]$ the probability that an order is executed where we assume for simplicity that the probability of order execution is independent of the order size. Pricing in the DP is linear. Namely, for a given execution price π, the utility that a trader of type θ extracts from submitting an order of q shares to be traded in the DP is

$$p\big[(2\theta\alpha - \pi)q - \alpha q^2\big] - \kappa,$$

where again we have normalized reservation utilities to zero. The problem of optimal submission to the DP for a θ-type trader is

$$\max_q \Big\{p\big[(2\theta\alpha - \pi)q - \alpha q^2\big]\Big\},$$

which yields the optimal submission level

$$q_d(\theta) := \theta - \frac{\pi}{2\alpha}.$$

We obtain that opting for the DP results in a trader of type θ enjoying the expected utility

$$u_0(\theta; \pi) = \alpha p \left(\theta - \frac{\pi}{2\alpha}\right)^2 - \kappa.$$

We assume that $p\pi^2 < 4\alpha\kappa$ so as to keep the DP unattractive for small types.

We assume that the dealer's costs/profits of unwinding a portfolio of size q are $C(q) = \epsilon\, q + \beta q^2$ where $\beta > 0$ and ϵ is non-negative. Observe that, since $u_0(\cdot;\pi)$ does not satisfy Assumption 1.2.2, some restrictions must be imposed on the problem's parameters so as to still have Lemma 1.5.1. Namely, it must hold that

$$\pi < 2\sqrt{\frac{\alpha\kappa}{p}}. \tag{1.6.1}$$

Condition (1.6.1) imposes a hard upper bound on possible equilibrium DP prices. It should be noted that Assumption 1.2.8 is not satisfied by $u_0(\cdot;\pi)$, which, together with the way in which we shall define the pricing feedback loop from the DM to the DP, implies that our equilibrium result does not apply "as is" to the current setting.

1.6.1 The dealer market without a dark pool

In the absence of a DP, the dealer's optimal choices of quantities are

$$l(\theta, 0) = \frac{\alpha}{\alpha+\beta}\big(2\theta + 1\big) - \frac{\epsilon}{2(\alpha+\beta)}$$

for negative types and

$$l(\theta, 1) = \frac{\alpha}{\alpha+\beta}\big(2\theta - 1\big) - \frac{\epsilon}{2(\alpha+\beta)}$$

for positive types, where the boundary of Θ_0 is given by

$$\underline{\theta}_0 = \frac{1}{2}\Big(\frac{\epsilon}{2\alpha} - 1\Big) \quad \text{and} \quad \overline{\theta}_0 = \frac{1}{2}\Big(\frac{\epsilon}{2\alpha} + 1\Big).$$

In order to guarantee that $\Theta_0 \subset [-1, 1]$ the condition $\epsilon < 2\alpha$ must be imposed on the corresponding parameters. From the relation $v'(\theta) = \psi_1\big(q(\theta)\big)$ we have that the indirect–utility function is

$$v(\theta) = \begin{cases} \frac{2\alpha^2}{\alpha+\beta}\theta^2 + \frac{\alpha}{\alpha+\beta}\big(2\alpha - \epsilon\big)\theta + c_1, & \theta \le \underline{\theta}_0; \\ \frac{2\alpha^2}{\alpha+\beta}\theta^2 - \frac{\alpha}{\alpha+\beta}\big(2\alpha + \epsilon\big)\theta + c_2, & \theta \ge \overline{\theta}_0, \end{cases}$$

where

$$c_1 = \frac{2\alpha^2}{4(\alpha+\beta)}\Big(\frac{\epsilon}{2\alpha} + 1\Big)^2 + \frac{\alpha(2\alpha+\epsilon)}{2(\alpha+\beta)}\Big(\frac{\epsilon}{2\alpha} + 1\Big) \quad \text{and}$$

$$c_2 = \frac{2\alpha^2}{4(\alpha+\beta)}\Big(\frac{\epsilon}{2\alpha} - 1\Big)^2 - \frac{\alpha(2\alpha+\epsilon)}{2(\alpha+\beta)}\Big(\frac{\epsilon}{2\alpha} - 1\Big).$$

When it comes to the spread, observe that $q' \equiv \frac{2\alpha}{\alpha+\beta}$, $\psi_1 \equiv 2\alpha$ and $\psi_2 \equiv 0$, which yields

$$[t(0_-), t(0_+)] = \frac{4\alpha^2}{\alpha+\beta}[\underline{\theta}_0, \overline{\theta}_0].$$

Below we analyze how the spread changes with the introduction of the DP.

1.6.2 The impact of a dark pool

We first take an exogenous execution price π and determine, for each $\theta \in \Theta$, what is the quantity–price pair $\big(q_c(\theta;\pi), \tau_c(\theta;\pi)\big)$ that the dealer must offer so as to match a DP with execution price π. Using the relation $q_c(\theta;\pi) = u_0'(\theta;\pi)$ we obtain

$$\begin{aligned} &q_c(\theta;\pi) = 2\alpha p\left(\theta - \frac{\pi}{2\alpha}\right) \text{ and} \\ &\tau_c(\theta;\pi) = \kappa + 4\alpha^2 p(\theta - \alpha p)\left(\theta - \frac{\pi}{2\alpha}\right) - \alpha p\left(\theta - \frac{\pi}{2\alpha}\right)^2. \end{aligned} \tag{1.6.2}$$

From the Envelope Theorem and the structure of $u(\theta, q)$ we have that the traders' indirect–utility function satisfies

$$\frac{v'(\theta)}{2\alpha} = l\big(\theta, \gamma(\theta)\big). \tag{1.6.3}$$

In order to determine the spread in the presence of the DP we must determine $\underline{\theta}_{0,m}$ and $\overline{\theta}_{0,m}$ together with $\gamma\big(\underline{\theta}_{0,m}\big)$ and $\gamma\big(\overline{\theta}_{0,m}\big)$. For an arbitrary $\Gamma \in [0,1]$ we have

$$l(\theta, \Gamma) = \frac{\alpha}{\alpha+\beta}\big[2\theta + 1 - 2\Gamma\big] - \frac{\epsilon}{2(\alpha+\beta)}.$$

Indexed by Γ, the candidates for $\underline{\theta}_{0,m}$ are then given by

$$\underline{\theta}_{0,m}(\Gamma) = \frac{1}{2}\Big(\frac{\epsilon}{2\alpha} + 2\Gamma - 1\Big).$$

Since it must hold that $\underline{\theta}_{0,m}(\Gamma) \leq 0$, then $\Gamma \leq 0.5(1 - \epsilon/2\alpha)$. Integrating Expression (1.6.3) we have that, on the interval $[\widetilde{\theta}_m(\Gamma), \underline{\theta}_{0,m}(\Gamma)]$, the traders' indirect utility is given by

$$v(\theta;\Gamma) = \frac{2\alpha^2}{\alpha+\beta}\theta^2 + 2\alpha\Big[\frac{\alpha}{\alpha+\beta}(1-2\Gamma) - \frac{\epsilon}{2(\alpha+\beta)}\Big]\theta + c_{1,m}, \tag{1.6.4}$$

where $\widetilde{\theta}_m(\Gamma)$ is the first intersection to the left of $\underline{\theta}_{0,m}(\Gamma)$ of $v(\cdot;\Gamma)$ and $u_0(\cdot;\pi)$ and $c_{1,m}$ is determined by the equation

$$v\big(\underline{\theta}_{0,m}(\Gamma);\Gamma\big) = 0.$$

Unless the inequality $\Gamma \leq 0.5(1-\epsilon/2\alpha)$ is tight, in which case the types below $\widetilde{\theta}_m(\Gamma)$ are excluded, Proposition 1.4.5 implies that Γ must be chosen so as to satisfy the smooth–pasting condition $u_0'\big(\widetilde{\theta}_m(\Gamma);\pi\big) = v'\big(\widetilde{\theta}_m(\Gamma);\pi\big)$, which is equivalent to

$$\widetilde{\theta}_m(\Gamma) = \Big[\frac{2\alpha}{\alpha+\beta} - p\Big]^{-1}\Big[\frac{\epsilon}{2(\alpha+\beta)} - \frac{\alpha}{\alpha+\beta}(1-2\Gamma) - \frac{p\pi}{2\alpha}\Big].$$

Observe that, besides the requirement $\Gamma \geq 0.5(1-\epsilon/2\alpha)$, the strategy to determine $\overline{\theta}_{0,m}$ is exactly the same as for $\underline{\theta}_{0,m}$. Summarizing, from Eq. (1.6.4) we observe that, if Γ_- and Γ_+ correspond to the optimal choices for the negative and positive endpoints of $\Theta_0(\pi)$, then

$$q'\big(\underline{\theta}_{0,m}(\Gamma_-)\big) = \frac{1}{2\alpha}v''\big(\underline{\theta}_{0,m}(\Gamma_-);\Gamma_-\big) = \frac{1}{2\alpha}v''\big(\overline{\theta}_{0,m}(\Gamma_+);\Gamma_+\big) = q'\big(\overline{\theta}_{0,m}(\Gamma_+)\big) = \frac{2\alpha}{\alpha+\beta}.$$

The spread is then

$$[t_m(0_-), t_m(0_+)] = \frac{4\alpha^2}{\alpha+\beta}[\underline{\theta}_{0,m}(\Gamma_-), \overline{\theta}_{0,m}(\Gamma_+)] \subset \frac{4\alpha^2}{\alpha+\beta}[\underline{\theta}_0, \overline{\theta}_0],$$

i.e. the presence of a DP strictly narrows the spread in the DM.

1.6.3 An equilibrium price

A standard (but not unique) way in which DP prices are generated is by computing the average of some publicly available best bid and best ask prices. In the case of the US, this is usually the mid-quote of the National Best Bid and Offer (NBBO). Borrowing from this idea we define the price iteration in the DP as follows:

$$\pi_{i+1} = \frac{1}{2}\big(t_i(0_+) - t_i(0_-)\big), \quad i \in \mathbb{N},$$

where $\{t_i(0_-), t_i(0_+)\}$ are the best bid and ask prices in the DM in the presence of a DP with execution price π_i. We know from the previous section that the sequence $\{\pi_i, i \in \mathbb{N}\} \subset ((4\alpha^2)/(\alpha+\beta))[\underline{\theta}_0, \overline{\theta}_0]$; hence, by the Bolzano-Weierstrass Theorem it has at least one convergent subsequence. The limit of each of the said subsequences will be an equilibrium price. The (possible) non-uniqueness of these prices is due to the fact that by virtue of its definition, the sequence of DP prices need not be monotonic. The problem of non-uniqueness of equilibria in models of competing DMs and CNs has been observed before. We refer to [DDH13] for a detailed discussion.

1.7 Conclusions

We have presented a screening model to study the structure of the LOB of a dealer who provides liquidity to traders of unknown preferences. Furthermore, we have established a link between the traders' indirect–utility function and the bid–ask spread in the DM. Making use of the aforementioned link, we have studied how the presence of a type-dependent outside option impacts the spread of the DM, as well as the set of trader types who participate in the DM and their welfare. In particular, we have shown, in a portfolio-liquidation setting, that the presence of a dark pool results in a shrinkage of the spread in the DM. Finally, we have established that, under certain conditions, the feedback loop introduced by the impact that the spread has on the structure of the outside option leads to an equilibrium price.

2 | Equilibrium Pricing Under Relative Performance Concerns

2.1 Organization of this chapter

In Section 2.3, we define the general market, agents, optimization problem and equilibrium that we consider. Sections 2.4 and 2.5 are devoted to solving the general optimization problem for a set of agents having arbitrary risk measures. In the former we solve the optimization problem for each agent, given the strategies of all others, i.e., we solve the best response problem, and we find a unique NE. In the latter we characterize the equilibrium price by means of a suitable weighted aggregation of individual risk measures and identification of the representative agent. Sections 2.6 and 2.7 analyze the particular case where the agents use entropic risk measures. In this more tractable setting, Section 2.6 explores theoretically the influence of various parameters on the global risk while Section 2.7 focuses on a model with two agents with opposite risk profiles, and numerically explores the influence of the concern rates, in particular, on the individual behaviors, risks, and the consequences for the whole system. Section 2.8 concludes the chapter.

2.2 Spaces and Notation

We define the following spaces for $p > 1$, $q \geq 1$, $n, m, d, k \in \mathbb{N}$: $C^{0,n}([0,T] \times \mathbb{R}^d; \mathbb{R}^k)$ is the space of continuous functions endowed with the $\|\cdot\|_\infty$-norm that are n-times continuously differentiable in the spatial variable; $C_b^{0,n}$ contains all bounded functions of $C^{0,n}$; the first superscript 0 is dropped for functions independent of time; $L^p(\mathcal{F}_t, \mathbb{R}^d)$, $t \in [0,T]$, is the space of d-dimensional $\mathcal{F}_t$-measurable random variables X with norm $\|X\|_{L^p} = \mathbb{E}[\,|X|^p]^{1/p} < \infty$; L^∞ refers to the subset of essentially bounded random variables; $\mathcal{S}^p([0,T] \times \mathbb{R}^d)$ is the space of d-dimensional measurable $\mathcal{F}$-adapted processes Y satisfying $\|Y\|_{\mathcal{S}^p} = \mathbb{E}[\sup_{t\in[0,T]} |Y_t|^p]^{1/p} < \infty$; $\mathcal{S}^\infty$ refers to the subset of $\mathcal{S}^p(\mathbb{R}^d)$ of essentially bounded processes; $\mathcal{H}^p([0,T] \times \mathbb{R}^d)$ is the space of d-dimensional measurable $\mathcal{F}$-adapted processes Z satisfying $\|Z\|_{\mathcal{H}^p} = \mathbb{E}[\big(\int_0^T |Z_s|^2 \mathrm{d}s\big)^{p/2}]^{1/p} < \infty$; For a probability measure $\mathbb{Q}$, we denote $\mathcal{H}_{\text{BMO}}(\mathbb{Q})$ as the space of processes $Z \in \mathcal{H}^p(\mathbb{Q})$ for any $p \geq 2$ such that for some constant $K_{BMO} > 0$

$$\sup_{\tau \in \mathcal{T}_{[0,T]}} \mathbb{E}^{\mathbb{Q}}\left[\int_\tau^T |Z_s|^2 \mathrm{d}s \Big| \mathcal{F}_\tau\right] \leq K_{BMO} < \infty,$$

where $\mathcal{T}_{[0,T]}$ is the set of all stopping times $\tau \in [0,T]$. For the reference measure $\mathbb{P}$ we write directly $\mathcal{H}_{\text{BMO}}$ instead of $\mathcal{H}_{\text{BMO}}(\mathbb{P})$. For more properties of $\mathcal{H}_{\text{BMO}}$, see Appendix A.1.

2.3 The model

We consider a finite set $\mathbb{A}$ of N agents, without loss of generality $\mathbb{A} = \{1, 2, \dots, N\}$, with random endowments H^a, $a \in \mathbb{A}$, to be received at a terminal time $T < \infty$. They trade continuously in the financial market which comprises a stock and a newly introduced structured security (called derivative), aiming to minimize their risk. For simplicity we assume that money can be lent or borrowed at the risk-free rate zero. Stock prices follow an *exogenous* diffusion process and are not affected by the agents' demand. By contrast, the derivative is traded only by the agents from $\mathbb{A}$ and priced *endogenously* such that demand matches supply.
The market model follows [HPDR10], whereas the preferences are extended by a performance functional as in [ET15] and [FdR11].

2.3.1 The market

Sources of risk and underlyings

Throughout this chapter we work on a continuous time scale $t \in [0,T]$. In our model, there are two independent sources of randomness, represented by a 2-dimensional standard Brownian motion $W = (W^S, W^R)$ on a standard filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t=0}^T, \mathbb{P})$, where $(\mathcal{F}_t)$ is the filtration generated by W and augmented by the $\mathbb{P}$-null sets and $\mathcal{F} = \mathcal{F}_T$. The Brownian motion W^R drives the external and non-tradable risk process (R_t), which is thought of as a temperature process or a precipitation index. For analytical convenience we assume that (R_t) follows a Brownian motion with drift being a stochastic process μ^R and constant volatility $b > 0$, i.e.,

$$\mathrm{d}R_t = \mu_t^R \mathrm{d}t + b\mathrm{d}W_t^R, \qquad \text{with } R_0 = r_0 \in \mathbb{R}. \tag{2.3.1}$$

The Brownian motion W^S drives the stock price process (S_t) according to

$$\begin{aligned} \mathrm{d}S_t &= \mu_t^S S_t \, \mathrm{d}t + \sigma_t^S S_t \, \mathrm{d}W_t^S, \\ &= \mu_t^S S_t \, \mathrm{d}t + \langle \sigma_t, \mathrm{d}W_t \rangle \quad \text{with } \sigma_t := (\sigma_t^S S_t, 0) \in \mathbb{R}^2, \text{ and } S_0 = s_0 > 0. \end{aligned} \tag{2.3.2}$$

We assume that the stochastic processes $\mu^R, \mu^S, \sigma^S : \Omega \times [0,T] \to \mathbb{R}$ are $(\mathcal{F}_t)$-adapted, with $\sigma^S > 0$.

Market price of risk: financial and external

We recall (see e.g. [HM07]) that any linear pricing scheme on the set $L^2(\mathbb{P})$ of square-integrable random variables with respect to $\mathbb{P}$ can be identified with a 2-dimensional predictable process $\theta = (\theta^S, \theta^R)$ such that the exponential process $(\mathcal{E}_t^\theta)$ defined by

$$\mathcal{E}_t^\theta := \mathcal{E}\left(-\int_0^\cdot \langle \theta_s, \mathrm{d}W_s \rangle\right)_t = \exp\left\{-\int_0^t \langle \theta_s, \mathrm{d}W_s \rangle - \frac{1}{2}\int_0^t |\theta_s|^2 \mathrm{d}s\right\}, \quad t \in [0,T], \tag{2.3.3}$$

is a uniformly integrable martingale. This ensures that the measure $\mathbb{P}^\theta$ defined by having density $\mathcal{E}_T^\theta$ against $\mathbb{P}$ is indeed a probability measure (the *pricing measure*), and the present price of a random terminal payment X is given by $\mathbb{E}^\theta[X]$, where $\mathbb{E}^\theta$ denotes the expectation with respect to $\mathbb{P}^\theta$. For any such θ, we introduce the $\mathbb{P}^\theta$-Brownian motion

$$W_t^\theta = W_t + \int_0^t \theta_s \, \mathrm{d}s, \qquad t \in [0, T]. \tag{2.3.4}$$

The first component θ^S of the vector θ is the *market price of financial risk*. Under the assumption that there is no arbitrage, S must be a martingale under $\mathbb{P}^\theta$ and, from the exogenously given dynamics of S, θ^S is necessarily given by $\theta_t^S = \mu_t^S/\sigma_t^S$, $t \in [0, T]$. The process θ^R on the other hand is unknown. It is the *market price of external risk* and will be derived endogenously by the market clearing condition (or constant net supply condition, see below).

The agents' endowments and the derivative's payoffs

The agents $a \in \mathbb{A}$ receive at time T the income H^a which depends on the financial and external risk factors. While the agents are able to trade in the financial market to hedge away some of their financial risk, a basis risk remains, originating in the agents' exposure to the non-tradable risk process R. A derivative with payoff H^D at maturity time T is introduced such that, by trading in the derivative H^D, the agents have now a way to reduce their basis risk.
We make the following assumptions on the endowments, derivative payoff and coefficients appearing in the dynamics of S and R.

Assumption 2.3.1 (Standing assumption on the data). *The processes μ^R, μ^S, σ^S and $\theta^S := \mu^S/\sigma^S$ are bounded (belong to $\mathcal{S}^\infty$). The random variables H^D and H^a, $a \in \mathbb{A}$, are bounded (belong to $L^\infty(\mathcal{F}_T)$).*

Price of the derivative, trading in the market and the agent's strategies

Assuming no arbitrage opportunities, the price process $(B_t^\theta)_{t\in[0,T]}$ of H^D is given by its expected payoff under $\mathbb{P}^\theta$; in other words $B_\cdot^\theta = \mathbb{E}^\theta\left[H^D|\mathcal{F}_\cdot\right]$. Since H^D is bounded, writing the $\mathbb{P}^\theta$-martingale as a stochastic integral against the $\mathbb{P}^\theta$-Brownian motion W^θ (with the martingale representation theorem) yields a 2-dimensional square-integrable adapted process $\kappa^\theta := (\kappa^S, \kappa^R)$ such that for $t \in [0, T]$

$$B_t^\theta = \mathbb{E}^\theta[H^D] + \int_0^t \langle \kappa_s^\theta, \mathrm{d}W_s^\theta\rangle = \mathbb{E}^\theta[H^D] + \int_0^t \langle \kappa_s^\theta, \mathrm{d}W_s\rangle + \int_0^t \left\langle \kappa_s^\theta, \theta_s\right\rangle \mathrm{d}s. \tag{2.3.5}$$

Note that we have $(B, \kappa) \in \mathcal{S}^\infty \times \mathcal{H}_{\mathrm{BMO}}(\mathbb{P}^\theta)$. It will turn out to be useful to rewrite (2.3.5) as a BSDE:

$$B_t^\theta = H^D - \int_t^T \langle \kappa_s^\theta, \theta_s\rangle \mathrm{d}s - \int_t^T \langle \kappa_s^\theta, \mathrm{d}W_s\rangle. \tag{2.3.6}$$

We denote by $\pi_t^{a,1}$ and $\pi_t^{a,2}$ the number of units agent $a \in \mathbb{A}$ holds in the stock and the derivative at time $t \in [0, T]$, respectively. Using a self-financing strategy $\pi^a := (\pi^{a,1}, \pi^{a,2})$

with values in $\mathbb{R}^2$, her gains from trading up to time $t \in [0,T]$, under the pricing measure $\mathbb{P}^\theta$ inducing the prices (B_t^θ) for the derivative, are given by

$$\begin{aligned} V_t^a = V_t(\pi^a) &= \int_0^t \pi_s^{a,1}\, \mathrm{d}S_s + \int_0^t \pi_s^{a,2}\, \mathrm{d}B_s^\theta \\ &= \int_0^t \left\langle \pi_s^{a,1}\sigma_s + \pi_s^{a,2}\kappa_s^\theta, \theta_s \right\rangle\, \mathrm{d}s + \int_0^t \langle \pi_s^{a,1}\sigma_s + \pi_s^{a,2}\kappa_s^\theta, \mathrm{d}W_s\rangle. \end{aligned} \tag{2.3.7}$$

We require that the trading strategies be integrable against the prices, i.e., for all $a \in \mathbb{A}$, $\pi^a \in L^2\big((S, B^\theta), \mathbb{P}^\theta\big)$, so that the gains processes are square-integrable martingales under $\mathbb{P}^\theta$ (i.e., we require $\mathbb{E}^\theta\left[\langle\, V_\cdot(\pi^a)\, \rangle_T\right] < \infty$). The $(\mathbb{R}^2)^{(N-1)}$-valued vector of strategies of all agents $b \in \mathbb{A} \setminus \{a\}$ will be denoted by $\pi^{-a} := (\pi^b)_{b \in \mathbb{A}\setminus\{a\}}$.

2.3.2 Preferences, risk minimization and equilibrium

The agents' measure of risk

The agents assess their risk using a dynamic convex time-consistent risk measure $\rho_\cdot^a$ induced by a BSDE.This means that the risk $\rho_t^a(\xi^a)$ which agent $a \in \mathbb{A}$ associates at time $t \in [0,T]$ with an $\mathcal{F}_T$-measurable random position ξ^a is given by Y_t^a, where (Y^a, Z^a) is the solution to the BSDE

$$-\mathrm{d}Y_t^a = g^a(t, Z_t^a)\mathrm{d}t - \langle Z_t^a, \mathrm{d}W_t\rangle \quad \text{with terminal condition} \quad Y_T^a = -\xi^a.$$

The driver g^a encodes the risk preferences of agent a for $a \in \mathbb{A}$. We assume that g^a has the following properties:

Assumption 2.3.2. *The map $g^a : [0,T] \times \mathbb{R}^2 \to \mathbb{R}$ is a deterministic continuous function. Its restriction to the space variable, $z \mapsto g^a(\cdot, z)$, is continuously differentiable, strictly convex and attains its minimum.*

For any fixed $(t, \vartheta) \in [0,T] \times \mathbb{R}^2$, the map $z \mapsto g^a(t,z) - \langle z, \vartheta\rangle$ is also strictly convex and attains its unique minimum at the point where its gradient vanishes. With this in mind we can define $\mathcal{Z}^a : [0,T] \times \mathbb{R}^2 \to \mathbb{R}^2$, $(t,\vartheta) \mapsto \mathcal{Z}^a(t,\vartheta)$ where $\mathcal{Z}^a(t,\vartheta)$ is the unique solution, in the unknown $\mathcal{Z}$, to the equation[1]

$$\nabla_z g^a(t, \mathcal{Z}) = \vartheta. \tag{2.3.8}$$

Under Assumption 2.3.2, the risk measure given by the above BSDE is strongly time consistent, convex and translation invariant (or monetary). For more details on the class of risk measures described by BSDEs we point the interested reader to [BE05], [Gia06] and [BE09].

For convenience, we recall the relevant properties of dynamic risk measures:

translation invariance: for any $m \in \mathbb{R}$ and any $t \in [0,T]$ it holds that $\rho_t^a(\xi^a + m) = \rho_t^a(\xi^a) - m$;

time-consistency: for any $t, t+s \in [0,T]$ it holds that $\rho_t^a(\xi^a) = \rho_t^a(\rho_{t+s}^a(\xi^a))$;

convexity: for any $t \in [0,T]$ and for $\xi^a, \hat{\xi}^a$ $\mathcal{F}_T$-measurable and $\alpha \in [0,1]$ we have $\rho_t^a\big(\alpha\xi^a + (1-\alpha)\hat{\xi}^a\big) \le \alpha\rho_t^a(\xi^a) + (1-\alpha)\rho_t^a(\hat{\xi}^a)$.

[1] We write $\nabla_z g^a(t, \mathcal{Z})$ for the vector consisting of the partial derivatives of g^a w.r.t. the (two) components of the space variable, evaluated at $\mathcal{Z}$.

The individual optimization problem

Agent a's position at maturity, ξ^a, is given by the sum of her terminal income H^a and the trading gains V_T^a over the time period $[0,T]$. Additionally, the agent compares her trading gains $V_T^a = V_T(\pi^a)$ with the average gains of all other agents. Thus, we define the perceived total wealth $\xi^a(\pi^a, \pi^{-a})$ of each of the N agents $a \in \mathbb{A}$ in the market at time $t = T$ as

$$\begin{aligned}\xi^a &= \Big(H^a + \big(1-\lambda^a\big)V_T(\pi^a)\Big) + \lambda^a\Big(V_T(\pi^a) - \frac{1}{N-1}\sum_{b\in\mathbb{A}\setminus\{a\}} V_T(\pi^b)\Big)\\ &= H^a + V_T(\pi^a) - \widetilde{\lambda}^a \sum_{b\in\mathbb{A}\setminus\{a\}} V_T(\pi^b), \quad \text{where} \quad \widetilde{\lambda}^a := \frac{\lambda^a}{N-1}\end{aligned}$$

and $\lambda^a \in [0,1]$ is the *concern rate* (or *jealousy factor*) of agent $a \in \mathbb{A}$ (cf. (0.0.1)).[2] We make the following assumption on the concern rates $\lambda^{\cdot}$, whose justification will become clear later on in Theorem 2.4.5.

Assumption 2.3.3 (Performance concern rates). *We have $\lambda^a \in [0,1]$ for each agent and $\prod_{a\in\mathbb{A}} \lambda^a < 1$.*

For notational convenience we introduce for $t \in [0,T]$

$$\bar{V}_t^{-a} := V_t(\bar{\pi}^{-a}) = \sum_{b\in\mathbb{A}\setminus\{a\}} V_t(\pi^b), \qquad \text{with} \qquad \bar{\pi}^{-a} := \sum_{b\in\mathbb{A}\setminus\{a\}} \pi^b, \tag{2.3.9}$$

where we apply the additivity of $\pi \mapsto V_{\cdot}(\pi)$, which can be inferred from its definition in (2.3.7). The risk associated with the self-financing strategy π^a evolves according to the BSDE

$$\begin{aligned}-\mathrm{d}Y_t^a &= g^a\left(t, Z_t^a\right)\mathrm{d}t - \langle Z_t^a, \mathrm{d}W_t\rangle, \quad t \in [0,T]\\ Y_T^a &= -\xi^a(\pi^a,\pi^{-a}) = -\Big(H^a + V_T(\pi^a) - \widetilde{\lambda}^a V_T(\bar{\pi}^{-a})\Big).\end{aligned} \tag{2.3.10}$$

Now, we introduce a notion of admissibility for our problem.

Definition 2.3.4 (Admissibility). *Let $a \in \mathbb{A}$, and $\pi^{-a} = (\pi^b)_{b\in\mathbb{A}\setminus\{a\}}$ be integrable strategies for the other agents. The $\mathbb{R}^2$-valued strategy process π^a is called* admissible *with respect to the market price of risk θ if $\mathbb{E}^\theta\left[\langle\, V_{\cdot}(\pi^a)\, \rangle_T\right] < \infty$, where $\langle V_{\cdot}(\pi^a)\rangle$ denotes the quadratic variation of $\big(V_t(\pi^a)\big)_{t\in[0,T]}$, and BSDE* (2.3.10) *has a unique solution. The set of admissible trading strategies for agent $a \in \mathbb{A}$ is denoted by $\mathcal{A}^\theta(\pi^{-a})$.*

Each agent $a \in \mathbb{A}$ wants to minimize her risk given the strategies of the other agents, π^{-a}, i.e. agent a solves the best-response problem

$$\min_{\pi^a\in\mathcal{A}^\theta(\pi^{-a})} Y_0^a(\pi^a,\pi^{-a}). \tag{2.3.11}$$

Notice that, a priori, not only the strategy chosen, but also the risk for agent a depends on the strategies of all other players, π^{-a}. For the sake of presentation we leave this interdependence implicit whenever possible and we write the solution to the BSDE giving the risk for agent a as (Y^a, Z^a) instead of the lengthy $\big(Y^a(\pi^a,\pi^{-a}), Z^a(\pi^a,\pi^{-a})\big)$. We will use the latter when the situation requires it.

[2] Compare with the performance functionals in [ET15] or [FdR11]. If $\lambda^a = 0$ for all $a \in \mathbb{A}$, then we are in the setting of [HPDR10].

Competitive equilibrium, equilibrium market price of risk and endogenous trading

We denote by $n \in \mathbb{R}$ the number of units of derivative present in the market. While each unit of derivative pays H^D at time T, the agents are free to buy and underwrite contracts for any amount of H^D. Within the trading period $[0, T]$, only the agents in our set $\mathbb{A}$, with trading objectives as described above, are active in the market and the total number n of derivatives present is constant over time.

We show that one can convert the problem for a general $n \in \mathbb{R}$ into a problem for $n = 0$ by distributing the derivative among all agents before the beginning of the trading period. In the case $n = 0$ every derivative held by an agent has been underwritten by another agent in $\mathbb{A}$, entailing essentially that agents share their risks with each other (see [BE05], [BE09] or [HM07]).

We assume that each agent seeks to minimize her risk measure independently, without cooperation with the other agents, so we are interested in Nash equilibria.

Definition 2.3.5 (Equilibrium and Equilibrium MPR (EMPR))**.** *For a given Market Price of Risk (MPR) $\theta = (\theta^S, \theta^R)$, we call $\pi^* = (\pi^{*,a})_{a\in\mathbb{A}}$ an* equilibrium *if, for all $a \in \mathbb{A}$, $\pi^{*,a} \in \mathcal{A}^\theta(\pi^{*,-a})$ and*

for any admissible strategy π^a it holds that $Y_0^a(\pi^{*,a}, \pi^{*,-a}) \le Y_0^a(\pi^a, \pi^{*,-a})$,

i.e., individual optimality given the strategies of the other agents. We call θ Equilibrium Market Price of Risk (EMPR) *and θ^R* Equilibrium Market Price of external Risk (EMPeR) *if*

1. *$\theta = (\theta^S, \theta^R)$ makes $\mathbb{P}^\theta$ a true probability measure (equivalently, $\mathcal{E}^\theta$ from* (2.3.3) *is a uniformly integrable martingale);*

2. *there exists a unique equilibrium π^* for θ;*

3. *π^* satisfies the market clearing condition (or fixed supply condition) for the derivative H^D (where Leb denotes the Lebesgue measure):*

$$\sum_{a\in\mathbb{A}} \pi_t^{*,a,2} = \sum_{a\in\mathbb{A}} \pi_0^{*,a,2} = n \quad \mathbb{P} \otimes Leb - a.e.. \tag{2.3.12}$$

Our approach to finding the equilibrium is illustrated in Figure 2.1.

2.4 The single agent's optimization and unconstrained equilibrium

Finding an EMPR is essentially an optimization problem under the fixed-supply constraint. So, prior to looking whether such an EMPR exists (postponed to Section 2.5), we start by fixing an arbitrary MPR $\theta \in \mathcal{H}_{\text{BMO}}$ and find the NE given that MPR, without the fixed-supply constraint. To this end, we first analyze the optimal behavior of the individual agents given that the others have chosen their strategies (the so-called best response problem), and then solve all best response problems simultaneously, thereby obtaining the NE, which will turn out to be unique for any given MPR θ.

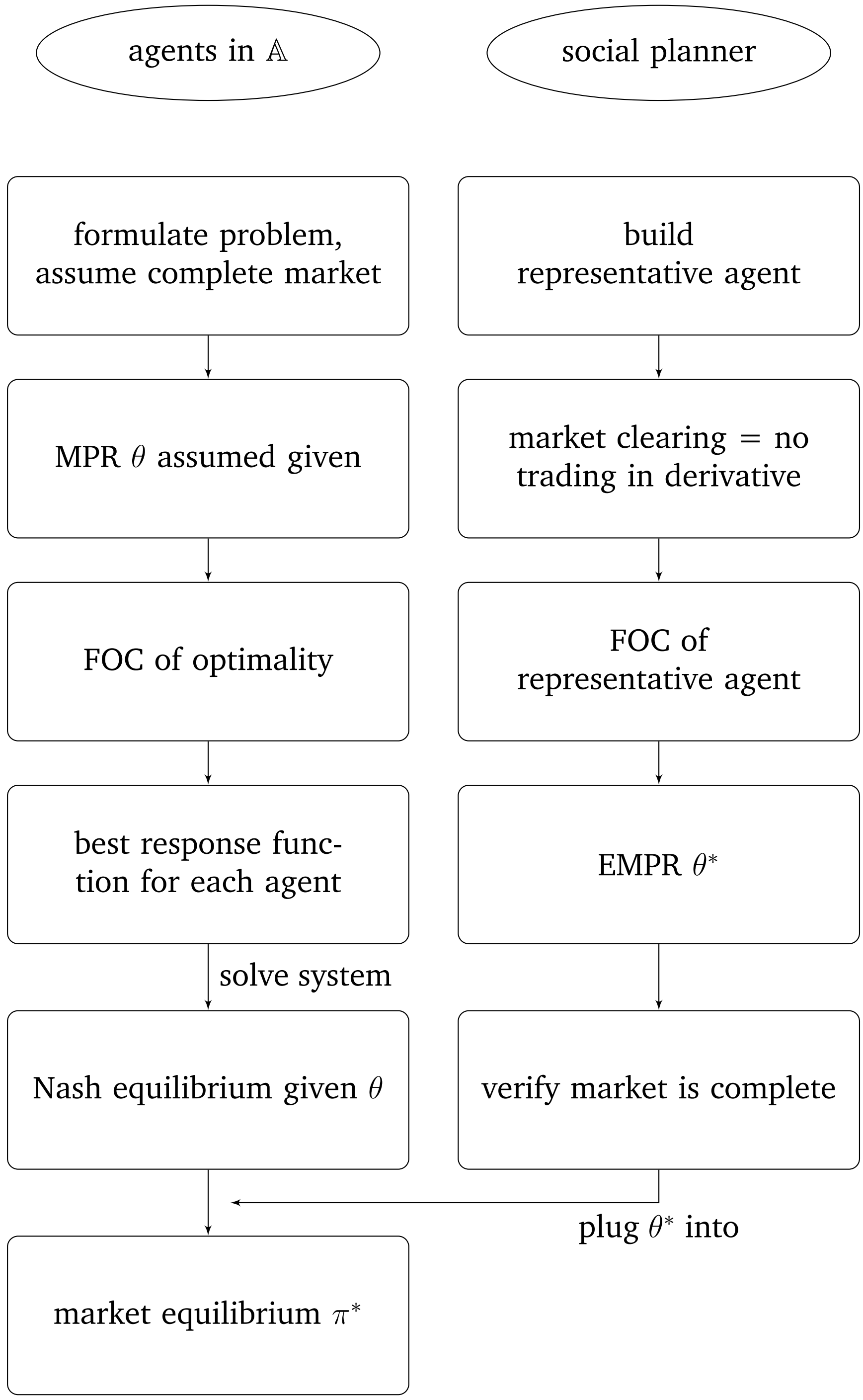


Figure 2.1: Stepwise Approach

2.4.1 Optimal response for one agent

In this subsection, in addition to a MPR θ being fixed, we focus on a single agent $a \in \mathbb{A}$, whose preferences are encoded by g^a, we take the strategies $\pi^{-a} = (\pi^b)_{b\in\mathbb{A}\setminus\{a\}}$ of the other agents as given, and we study the investment problem of our agent in this setting.

Optimizing the residual risk

To solve the optimization problem (2.3.11) for agent a, we first recall from [HPDR10] that, at each time $t \in [0,T]$, the strategy chosen must minimize the residual risk: the additivity of the risk measure implies (writing $V_T = (V_T - V_t) + V_t$ and using the translation invariance) that

$$\begin{aligned} Y_t^a &= \rho_t^a\Big(H^a + V_T^a - \widetilde{\lambda}^a \bar{V}_T^{-a}\Big) \\ &= \rho_t^a\Big(H^a + (V_T^a - V_t^a) - \widetilde{\lambda}^a(\bar{V}_T^{-a} - \bar{V}_t^{-a})\Big) - \big(V_t^a - \widetilde{\lambda}^a \bar{V}_t^{-a}\big). \end{aligned}$$

This suggests applying the following change of variables to (2.3.10) (using (2.3.9)),

$$\begin{aligned} \widetilde{Y}_t^a &:= Y_t^a + \big(V_t^a - \widetilde{\lambda}^a \bar{V}_t^{-a}\big), \\ \widetilde{Z}_t^a &:= Z_t^a + \zeta_t^a, \quad \text{where} \quad \zeta_t^a = \big(\pi_t^{a,1}\sigma_t + \pi_t^{a,2}\kappa_t^\theta\big) - \widetilde{\lambda}^a\big(\bar{\pi}_t^{-a,1}\sigma_t + \bar{\pi}_t^{-a,2}\kappa_t^\theta\big) \in \mathbb{R}^2. \end{aligned} \tag{2.4.1}$$

If the strategies are not clear from the context, we also write $\zeta^a = \zeta^a(\pi) = \zeta^a(\pi^a, \pi^{-a})$. Direct computations yield a BSDE for $(\widetilde{Y}^a, \widetilde{Z}^a)$ given by

$$\begin{aligned} -d\widetilde{Y}_t^a &= \widetilde{g}^a\Big(t, \pi_t^a, \pi_t^{-a}, \widetilde{Z}_t^a\Big)\mathrm{d}t - \langle \widetilde{Z}_t^a, \mathrm{d}W_t\rangle, \quad t \in [0,T], \\ \widetilde{Y}_T^a &= -H^a, \end{aligned} \tag{2.4.2}$$

where the driver $\widetilde{g}^a \colon \Omega \times [0,T] \times \mathbb{R}^2 \times (\mathbb{R}^2)^{N-1} \times \mathbb{R}^2 \to \mathbb{R}$ is defined as

$$\widetilde{g}^a(t, \pi_t^a, \pi_t^{-a}, z^a) : = g^a\big(t, z^a - \zeta_t^a\big) - \langle \zeta_t^a, \theta_t\rangle \tag{2.4.3}$$

$$= g^a\left(t, z^a - \Big(\big(\pi_t^{a,1} - \widetilde{\lambda}^a\bar{\pi}_t^{-a,1}\big)\sigma_t + \big(\pi_t^{a,2} - \widetilde{\lambda}^a\bar{\pi}_t^{-a,2}\big)\kappa_t^\theta\Big)\right) \tag{2.4.4}$$

$$- \Big\langle \big(\pi_t^{a,1} - \widetilde{\lambda}^a\bar{\pi}_t^{-a,1}\big)\sigma_t + \big(\pi_t^{a,2} - \widetilde{\lambda}^a\bar{\pi}_t^{-a,2}\big)\kappa_t^\theta, \theta_t \Big\rangle .$$

Each individual agent $a \in \mathbb{A}$ seeks to minimize $\widetilde{Y}_0^a$, the solution to (2.4.2), via her choice of investment strategy $\pi^a \in \mathcal{A}^\theta(\pi^{-a})$, in other words she aims at solving

$$\min_{\pi^a \in \mathcal{A}^\theta(\pi^{-a})} \widetilde{Y}_0^a(\pi^a, \pi^{-a}). \tag{2.4.5}$$

Before we solve the individual optimization problem, we make the assumption that the derivative H^D does indeed complete the market. This must then be verified a posteriori (once the solution is computed) and case-by-case depending on the specific model.

Assumption 2.4.1. *Assume that $\kappa_t^R \neq 0$, for any $t \in [0,T]$, $\mathbb{P}$-a.s. .*

The pointwise minimizer for the single agent's residual risk

In (2.4.2), the strategy π^a appears only in the driver $\widetilde{g}^a$, not in the terminal condition. The comparison theorem for BSDEs suggests that in order to minimize $\widetilde{Y}_0^a(\pi^a)$ over admissible strategies π^a, one needs only to minimize the driver function $\widetilde{g}^a$ over π_t^a, for each fixed ω, t, π_t^{-a} and z^a. We define such pointwise minimizer as the random map $\Pi^a\colon \Omega \times [0,T] \times (\mathbb{R}^2)^{N-1} \times \mathbb{R}^2 \to \mathbb{R}^2$, given by

$$\Pi^a(t, \pi_t^{-a}, z) := \arg\min_{\pi^a \in \mathbb{R}^2} \widetilde{g}^a(t, \pi^a, \pi_t^{-a}, z).$$

The pointwise minimization problem has, under Assumption 2.3.2, a unique minimizer, which is characterized by the first order condition (FOC) for $\widetilde{g}^a$, i.e.,

$$\nabla_{\pi^a} \widetilde{g}^a(t, \pi^a, \pi_t^{-a}, z^a) = 0.$$

Recall that $\sigma = (\sigma^S S, 0)$. Using (2.4.4), the FOC is equivalently written as

$$\begin{aligned}
\partial_{\pi^{a,1}} \widetilde{g}^a(t, \pi^a, \pi^{-a}, z^a) = 0 &\iff \big\langle (\nabla_z g^a)(t, z^a - \zeta^a), -\sigma \big\rangle - \langle \sigma, \theta \rangle = 0 \\
&\iff g^a_{z^1}(t, z^a - \zeta^a) = -\theta^S, && (2.4.6) \\
\partial_{\pi^{a,2}} \widetilde{g}^a(t, \pi^a, \pi^{-a}, z^a) = 0 &\iff \big\langle (\nabla_z g^a)(t, z^a - \zeta^a), -\kappa^\theta \big\rangle - \langle \kappa^\theta, \theta \rangle = 0 \\
&\iff -\theta^S \kappa^S + g^a_{z^2}(t, z^a - \zeta^a)\, \kappa^R = -\kappa^S \theta^S - \kappa^R \theta^R \\
&\iff g^a_{z^2}(t, z^a - \zeta^a) = -\theta^R, && (2.4.7)
\end{aligned}$$

where we used (2.4.6) to obtain (2.4.7) under Assumption 2.4.1.

With $\mathcal{Z}^a$ from (2.3.8), the FOC system (2.4.6)–(2.4.7) is equivalent to

$$z^a - \zeta_t^a = z^a - \zeta_t^a(\Pi^a(t, \pi_t^{-a}, z^a), \pi_t^{-a}) = \mathcal{Z}^a(t, -\theta_t), \tag{2.4.8}$$

which has the useful property that while the left-hand side (LHS) depends on z^a, the best response strategy of agent a and all other agents' strategies π^{-a}, the right-hand side (RHS) merely depends on the MPR θ and the structure of the driver g^a.

The expression for ζ^a in (2.4.1) and elementary re-arrangements allow to rewrite (2.4.8) as

$$\begin{aligned}
\Pi^{a,1}(t, \pi_t^{-a}, z^a) - \widetilde{\lambda}^a \bar{\pi}_t^{-a,1} &= \frac{z^{a,1} - \mathcal{Z}^{a,1}(t, -\theta_t)}{\sigma_t^S S_t} - \frac{z^{a,2} - \mathcal{Z}^{a,2}(t, -\theta_t)}{\kappa_t^R} \frac{\kappa_t^S}{\sigma_t^S S_t}, \\
\Pi^{a,2}(t, \pi_t^{-a}, z^a) - \widetilde{\lambda}^a \bar{\pi}_t^{-a,2} &= \frac{z^{a,2} - \mathcal{Z}^{a,2}(t, -\theta_t)}{\kappa_t^R}.
\end{aligned} \tag{2.4.9}$$

Plugging $z^a - \zeta_t^a = \mathcal{Z}^a(t, -\theta_t)$ into (2.4.3) yields an expression for the minimized (random) driver

$$\begin{aligned}
\widetilde{g}^a\big(t, \Pi^a(t, \pi_t^{-a}, z^a), \pi_t^{-a}, z^a\big) &= g^a\big(t, \mathcal{Z}^a(t, -\theta_t)\big) + \langle \mathcal{Z}^a(t, -\theta_t), \theta_t \rangle - \langle z^a, \theta_t \rangle \\
&=: \widetilde{G}^a(t, z^a).
\end{aligned} \tag{2.4.10}$$

We stress two important details. First, $\widetilde{G}^a$ is an affine driver (in z^a) with stochastic coefficients. Second, $\widetilde{G}^a$ does not depend at all on π^{-a} and it can be calculated without an explicit formula for the best response $\Pi^a(\cdot, \pi_\cdot^{-a}, z^a)$ by relying on $\mathcal{Z}^a$ from (2.3.8).

The second statement implies that, while the best-response strategy of an agent $a \in \mathbb{A}$ depends on the strategies of the other agents, her minimal risk does not.[3]

Both statements hold due to the affine structure of the agent's terminal position, which is passed on to ζ^a and via (2.4.3) to the driver, as well as due to Assumption 2.3.2 and (2.3.8), which lead to the replacement (2.4.8).

Single-agent optimality

Since $\widetilde{G}^a$ is an affine driver, and since $\nabla_z \widetilde{G}^a = -\theta \in \mathcal{H}_{\text{BMO}}$, we have a unique solution to the BSDE with driver $\widetilde{G}^a$ and terminal condition $-H^a$, provided that the process $(\omega, t) \mapsto \widetilde{G}^a(t, 0) = g^a\big(t, \mathcal{Z}^a(t, -\theta_t)\big) + \langle \mathcal{Z}^a(t, -\theta_t), \theta_t \rangle$ is sufficiently integrable. Let $(\widetilde{Y}^a, \widetilde{Z}^a)$ be the solution to BSDE (2.4.2) with driver (2.4.10) and define the strategy $\pi^{*,a}_\cdot := \Pi^a(\cdot, \pi^{-a}_\cdot, \widetilde{Z}^a_\cdot)$. This so-called best response function does not only solve the individual risk minimization problem, but, as we show next, it is even unique.

Theorem 2.4.2 (Optimality for one agent, uniqueness[4])**.** *Fix a market price of risk $\theta = (\theta^S, \theta^R) \in \mathcal{H}_{BMO}$ and let Assumption 2.4.1 hold. Fix an agent $a \in \mathbb{A}$ and a set of integrable strategies $(\pi^b)_{b \in \mathbb{A} \setminus \{a\}}$. Assume further that for $\widetilde{G}^a$ given by* (2.4.10)*, $|\widetilde{G}^a(\cdot, 0)|^{1/2} \in \mathcal{H}_{BMO}$. Then the BSDE with driver* (2.4.10) *and terminal condition $-H^a$ has a unique solution $(\widetilde{Y}^a, \widetilde{Z}^a) \in \mathcal{S}^\infty \times \mathcal{H}_{BMO}$. Moreover, if $\pi^{*,a}_\cdot = \Pi^a(\cdot, \pi^{-a}_\cdot, \widetilde{Z}^a_\cdot)$ is admissible, then $\widetilde{Y}^a_0$ is the value of the optimization problem* (2.4.5) *(i.e. the minimized risk) for agent a and $\pi^{*,a}$ is the unique optimal strategy.*

Proof. Given the structure of $\widetilde{G}^a$ in (2.4.10) and the integrability assumption made, the existence and uniqueness of the BSDE's solution $(\widetilde{Y}^a, \widetilde{Z}^a)$ in $\mathcal{S}^\infty \times \mathcal{H}_{\text{BMO}}$ is straightforward[5].

We first use the comparison theorem to prove the minimality of $\widetilde{Y}^a$, and hence the optimality of $\pi^{*,a}$. Let $t \in [0, T]$. Take any strategy $\pi^a \in \mathcal{A}^\theta(\pi^{-a})$. First, from the definition of $\widetilde{G}^a$ as a pointwise minimum, we naturally have that

$$\widetilde{G}^a(t, z^a) = \widetilde{g}^a\left(t, \Pi^a(t, \pi^{-a}_t, z^a), \pi^{-a}_t, z^a\right) \leq \widetilde{g}^a(t, \pi^a_t, \pi^{-a}_t, z^a) \text{ for all } t \in [0, T] \text{ and } z^a \in \mathbb{R}^2,$$

i.e., $\widetilde{G}^a(\cdot, \cdot) \leq \widetilde{g}^a(\cdot, \pi^a_\cdot, \pi^{-a}_\cdot, \cdot)$. Second, $\widetilde{G}^a$ is affine and thus Lipschitz continuous, with Lipschitz coefficient process $-\theta \in \mathcal{H}_{\text{BMO}}$. By the comparison theorem, we therefore have, for any $t \in [0, T]$ and in particular for $t = 0$, that $\widetilde{Y}^a_t = \widetilde{Y}^a_t(\pi^{*,a}, \pi^{-a}) \leq \widetilde{Y}^a_t(\pi^a, \pi^{-a})$. As this holds for any $\pi^a \in \mathcal{A}^\theta(\pi^{-a})$, this proves the minimality of $\widetilde{Y}^a_0 = \rho^a_0\big(\xi^a(\pi^{*,a}, \pi^{-a})\big)$ and thus the optimality of $\pi^{*,a}$.

We now argue the uniqueness of the optimizer $\pi^{*,a}$. Let π^a be an admissible strategy and let $(\widetilde{Y}^a(\pi^a), \widetilde{Z}^a(\pi^a))$ be the corresponding risk, i.e. solution to the BSDE (2.4.2)

[3] In [HPDR10] the driver did not depend on the other players' strategies in the first place. There is an equivalent assumption to Assumption 2.3.2 in [HPDR10, Proposition 3.6 (i)], but the observation of the linearity was not made.

[4] cf. [HPDR10, Proposition 3.6], which we extend by proving existence and uniqueness of the solution of the BSDE instead of assuming it. Uniqueness is important to us in order to obtain a unique NE for a given MPR θ.

[5] cf. [IDR10, Theorem 2.6], which states that $Y \in \mathcal{S}^\infty$ and $Z * W \in BMO$, which implies $Z \in \mathcal{H}_{\text{BMO}}$.

with strategy π^a. We compute the difference $\widetilde{Y}_t^a(\pi^a) - \widetilde{Y}_t^a(\pi^{*,a})$:

$$\begin{aligned}
&\widetilde{Y}_t^a(\pi^a) - \widetilde{Y}_t^a(\pi^{*,a}) \\
&= \int_t^T \left[\widetilde{g}^a\left(s, \pi_s^a, \pi_s^{-a}, \widetilde{Z}_s^a(\pi^a)\right) - \widetilde{G}^a\left(s, \widetilde{Z}_s^a(\pi^{*,a})\right)\right] \mathrm{d}s - \int_t^T [\widetilde{Z}_s^a(\pi^a) - \widetilde{Z}_s^a(\pi^{*,a})] \mathrm{d}W_s \\
&= \int_t^T \left[\widetilde{g}^a\left(s, \pi_s^a, \pi_s^{-a}, \widetilde{Z}_s^a(\pi^a)\right) - \widetilde{g}^a\left(s, \Pi^a\big(s, \pi_s^{-a}, \widetilde{Z}_s^a(\pi^a)\big), \pi_s^{-a}, \widetilde{Z}_s^a(\pi^a)\right)\right] \mathrm{d}s \qquad (2.4.11) \\
&\quad - \int_t^T [\widetilde{Z}_s^a(\pi^a) - \widetilde{Z}_s^a(\pi^{*,a})] \mathrm{d}W_s^\theta,
\end{aligned}$$

where we added $\widetilde{G}^a(t, \widetilde{Z}_t^a(\pi^a))$, subtracted $\widetilde{g}^a\left(t, \Pi^a\big(t, \pi_t^{-a}, \widetilde{Z}_t^a(\pi^a)\big), \pi_t^{-a}, \widetilde{Z}_t^a(\pi^a)\right)$ (equal to the added term) and used the affine structure[6] of $\widetilde{G}^a$ combined with (2.3.4).
By construction of Π^a as a minimizer, the difference in (2.4.11) is always positive. In particular, taking $\mathbb{P}^\theta$-expectation w.r.t. $\mathcal{F}_t$ implies that $\widetilde{Y}_t^a(\pi^a) - \widetilde{Y}_t^a(\pi^{*,a}) \geq 0$ for all $t \in [0, T]$. Assume that π^a is an optimal strategy. Then $\widetilde{Y}_0^a(\pi^a) = \widetilde{Y}_0^a(\pi^{*,a})$ and the LHS vanishes for $t = 0$. Under $\mathbb{P}^\theta$-expectation, the stochastic integral on the RHS also vanishes and we can conclude that the integrand in (2.4.11) is zero $\mathbb{P}^\theta \otimes Leb$-a.e.. Consequently, we obtain $\widetilde{Y}^a(\pi^a) = \widetilde{Y}^a(\pi^{*,a})$ and hence $\widetilde{Z}^a(\pi^a) = \widetilde{Z}^a(\pi^{*,a})$. This implies $\Pi^a\big(\cdot, \pi_\cdot^{-a}, \widetilde{Z}_\cdot^a(\pi^a)\big) = \Pi^a\big(\cdot, \pi_\cdot^{-a}, \widetilde{Z}_\cdot^a(\pi^{*,a})\big)$. Finally, by uniqueness of the minimizer Π^a, we obtain $\pi_\cdot^a = \Pi^a\big(\cdot, \pi_\cdot^{-a}, \widetilde{Z}_\cdot^a(\pi^a)\big) = \Pi^a\big(\cdot, \pi_\cdot^{-a}, \widetilde{Z}_\cdot^a(\pi^{*,a})\big) = \pi_\cdot^{*,a}$. □

Remark 2.4.3. *While Theorem 2.4.2 is stated as the optimal response of a single agent a in the system $\mathbb{A}$ with the other strategies π^{-a} being fixed, it is clear that it can more generally describe the optimal investment of an agent with preferences described by g^a (equivalently, $\rho_\cdot^a$) who trades in the assets S and B, which have the given MPR θ (one can think of setting $\mathbb{A} = \{a\}$ or $\lambda^a = 0$). Following the same methods, the result could be generalized to a higher number of assets, with price processes given exogenously. This applies similarly to an agent trading in fewer assets, by setting the respective components to zero – see Theorem 2.5.4.*

We now state a characterization of the optimal strategy via the FOC.

Lemma 2.4.4. *Under the assumptions of Theorem 2.4.2, let $\widehat{\pi}^a$ be an admissible strategy and $(\widehat{Y}^a, \widehat{Z}^a)$ be the associated risk process, solution to the BSDE with driver $\widetilde{g}^a(t, \widehat{\pi}_t^a, \pi_t^{-a}, \cdot)$ and terminal condition $-H^a$. Assume that they satisfy the FOC* (2.4.6)–(2.4.7) *in the sense that*

$$\nabla_z g^a(t, \widehat{Z}_t^a - \widehat{\zeta}_t^a) = -\theta_t \qquad \textit{where} \qquad \widehat{\zeta}_t^a = \big(\widehat{\pi}_t^{a,1}\sigma_t + \widehat{\pi}_t^{a,2}\kappa_t^\theta\big) - \widetilde{\lambda}^a\big(\bar{\pi}_t^{-a,1}\sigma_t + \bar{\pi}_t^{-a,2}\kappa_t^\theta\big).$$

Then $(\widehat{Y}^a, \widehat{Z}^a) = (\widetilde{Y}^a, \widetilde{Z}^a)$ and $\widehat{\pi}^a = \pi^{*,a}$.

Proof. By the assumptions on g^a, $\nabla_z g^a(t, \widehat{Z}_t^a - \widehat{\zeta}_t^a) = -\theta_t$ means that $\widehat{Z}_t^a - \widehat{\zeta}_t^a = \mathcal{Z}^a(t, -\theta_t)$, or equivalently $\widehat{\pi}_t = \Pi^a(t, \pi_t^{-a}, \widehat{Z}_t^a)$. Therefore, $\widetilde{g}^a(t, \widehat{\pi}_t^a, \pi_t^{-a}, \widehat{Z}_t^a) = \widetilde{G}^a(t, \widehat{Z}_t^a)$ – recall (2.4.3). By uniqueness of the solution to the BSDE with driver $\widetilde{G}^a(t, \cdot)$ and terminal condition $-H^a$, we have $(\widehat{Y}^a, \widehat{Z}^a) = (\widetilde{Y}^a, \widetilde{Z}^a)$. Consequently, by the uniqueness of the FOC's solution (Theorem 2.4.2), $\widehat{\pi}_t^a = \Pi^a(t, \pi_t^{-a}, \widehat{Z}_t^a) = \Pi^a(t, \pi_t^{-a}, \widetilde{Z}_t^a) = \pi_t^{*,a}$. □

[6] (2.4.10) implies $\widetilde{G}^a(t, z^a) - \widetilde{G}^a(t, \widehat{z}^a) = \langle \widehat{z}^a - z^a, \theta_t \rangle$ for any $t \in [0, T]$ and $z^a, \widehat{z}^a \in \mathbb{R}^2$.

2.4.2 The unconstrained Nash equilibrium

Having found the unique best response function of each agent, we now look at the existence and uniqueness of a NE, still for the MPR $\theta \in \mathcal{H}_{\text{BMO}}$ fixed at the beginning of this section, and still with no fixed-supply constraint.

Assume $\pi^* = (\pi^{*,a})_{a\in\mathbb{A}}$ is a NE. Fix an agent $a \in \mathbb{A}$. From the uniqueness of the optimal strategy, given by Theorem 2.4.2, one must have

$$\pi_t^{*,a} = \Pi^a(t, \pi_t^{*,-a}, \widetilde{Z}_t^a), \quad t \in [0,T],$$

where $(\widetilde{Y}^a, \widetilde{Z}^a)$ is the solution to the BSDE with terminal condition $-H^a$ and driver $\widetilde{G}^a$ given in (2.4.10). From the characterization (2.4.9) of Π^a, we therefore have, for all $a \in \mathbb{A}$ and $t \in [0,T]$,

$$\begin{aligned} \pi_t^{*,a,1} - \widetilde{\lambda}^a \bar{\pi}_t^{*,-a,1} &= \frac{\widetilde{Z}_t^{a,1} - \mathcal{Z}^{a,1}(t,-\theta_t)}{\sigma_t^S S_t} - \frac{\widetilde{Z}_t^{a,2} - \mathcal{Z}^{a,2}(t,-\theta_t)}{\kappa_t^R} \frac{\kappa_t^S}{\sigma_t^S S_t} =: J_t^{a,1}, \\ \pi_t^{*,a,2} - \widetilde{\lambda}^a \bar{\pi}_t^{*,-a,2} &= \frac{\widetilde{Z}_t^{a,2} - \mathcal{Z}^{a,2}(t,-\theta_t)}{\kappa_t^R} =: J_t^{a,2}. \end{aligned} \tag{2.4.12}$$

Note that, for any $a \in \mathbb{A}$, the process $(\widetilde{Y}^a, \widetilde{Z}^a)$ does not depend on π^*as neither $-H^a$ nor $\widetilde{G}^a$ does. Therefore, J_t^a is also independent of the unknown π_t^*, which is only present in the LHS of the System (2.4.12).

Conversely, assume we can solve for π^* in (2.4.12) and that π^* is integrable against the prices. Then, since $\pi_t^{*,a} = \Pi^a(t, \pi_t^{*,-a}, \widetilde{Z}_t^a)$ by (2.4.9), Theorem 2.4.2 guarantees that $\pi^{*,a}$ is the best response to $\pi^{*,-a}$, and we therefore have a NE.

Summing up, the existence and uniqueness of a NE π^* is equivalent to the existence and uniqueness of a solution to System (2.4.12).
Define the matrix $A_N \in \mathbb{R}^{N\times N}$ by[7]

$$A_N = \begin{pmatrix} 1 & & -\widetilde{\lambda}^1 \\ & \ddots & \\ -\widetilde{\lambda}^N & & 1 \end{pmatrix}, \quad \text{where } \widetilde{\lambda}^j = \frac{\lambda^j}{(N-1)}, \ j \in \mathbb{A}, \tag{2.4.13}$$

i.e., the j-th line has the entries $-\widetilde{\lambda}^j$ everywhere but in the j-th column, where it equals 1. System (2.4.12) can be rewritten as[8]

$$A_N \, \pi^{*,\cdot,i} = J^{\cdot,i}, \tag{2.4.14}$$

where $\pi^{*,\cdot,i} = (\pi^{*,a,i})_{a\in\mathbb{A}}$ and $J^{\cdot,i} = (J^{a,i})_{a\in\mathbb{A}}$, for $i \in \{1,2\}$.

Theorem 2.4.5. *Assume that the MPR $\theta = (\theta^S, \theta^R) \in \mathcal{H}_{BMO}$, that Assumption 2.4.1 and Assumption 2.3.3 hold, that, for all $a \in \mathbb{A}$, J^a is integrable against the prices and $|\widetilde{G}^a(\cdot,0)|^{1/2} \in \mathcal{H}_{BMO}$.*
Then there exists a unique NE $\pi^* = (\pi^{*,a})_{a\in\mathbb{A}}$ associated with the MPR θ, which is given by the unique solution to (2.4.14).

[7] Recall the notation that for sums and products over certain subsets of $\mathbb{A}$ we identify $\mathbb{A}$ with the set $\{1,2,\ldots,N\}$, where $N \in \mathbb{N}$ is the fixed finite number of agents.

[8] cf. [ET15, Section 3.2].

Proof. By [ET15, Remark 3.4], A_N is invertible if and only if $\prod_{a\in\mathbb{A}} \lambda^a < 1$, and under this condition the inverse is explicitly given. This guarantees that one can solve System (2.4.12) (or, equivalently, (2.4.14)) for each $i \in \{1,2\}$ to obtain $(\pi^{*,\cdot,i})$. The integrability of π^* follows from fact that each component $\pi^{*,a}$ is a linear combination of the integrable J^a's. Finally, the Nash-optimality of π^* was argued in the identification of Equation (2.4.12).[9] $\square$

Remark 2.4.6. *Without the explicit formula for A_N^{-1} one can also prove invertibility by looking at the determinant. The determinant of the A_N is (by Laplace's formula)*

$$\det(A_N) = 1 - \sum_{i<j} \widetilde{\lambda}^i\widetilde{\lambda}^j - 2\sum_{i<j<k} \widetilde{\lambda}^i\widetilde{\lambda}^j\widetilde{\lambda}^k - 3\sum_{i<j<k<l} \widetilde{\lambda}^i\widetilde{\lambda}^j\widetilde{\lambda}^k\widetilde{\lambda}^l - \ldots - (N-1)\prod_{i=1}^N \widetilde{\lambda}^i,$$

where the sums run over indices i,j,k,l from $\mathbb{A} = \{1,\ldots,N\}$. If $\lambda^a = 1$ for all $a \in \mathbb{A}$, then by Lemma A.4.1 $\det(A_N) = 1 - \sum_{k=2}^N \frac{k-1}{(N-1)^k}\binom{N}{k} = 0$, so the matrix is not invertible. The determinant is strictly decreasing in each $\widetilde{\lambda}^a$ ($a \in \mathbb{A}$) and therefore also in λ^a. Hence, if $\lambda^a \in [0,1]$ for all $a \in \mathbb{A}$ and the product $\prod_{a\in\mathbb{A}} \lambda^a < 1$, then at least one factor must be strictly smaller than one and the determinant must be strictly positive (i.e., $\det(A_N) > 0$); the invertibility of A_N follows.

We can now comment on Assumption 2.3.3. If $\lambda^b = 0$ for all $b \in \mathbb{A} \setminus \{a\}$, then A_N is invertible independently of λ^a, in particular for $\lambda^a = 1$. This shows that $\lambda^a \in [0,1)$ for all $a \in \mathbb{A}$ is not necessary, but merely a sufficient condition. Finally, if we were to allow for $\lambda^a > 1$, then $\prod_{a\in\mathbb{A}} \lambda^a < 1$ is not sufficient for invertibility of A_N, e.g. in the case $N = 3$ take $\lambda^a = \lambda^b = 2$ and $\lambda^c = 0$.
From now on we assume that the agents' optimization problems have a solution so that it makes sense to discuss the notion of EMPR.

Remark 2.4.7. *Notice that at this point, as θ is given exogenously, we have a system of uncoupled BSDEs. For each $a \in \mathbb{A}$, we obtain $(\widetilde{Y}^a, \widetilde{Z}^a)$ as the solution to a BSDE with terminal condition $-H^a$ and driver $\widetilde{G}^a$ from* (2.4.10)*, which only takes $\widetilde{Z}^a$ as argument, but does not depend on the strategies. These processes are then used to solve for the NE of strategies, π^*, given the MPR θ.*

2.4.3 An example: the entropic risk measure case

We now illustrate the methodology and result of Theorem 2.4.2 for a particular risk measure, and prepare the ground for the model we study in Sections 2.6 and 2.7. As before, a market price of risk $\theta = (\theta^S, \theta^R) \in \mathcal{H}_{\text{BMO}}$ is given (exogenously) and agents need not be concerned about market clearing.

Each agent $a \in \mathbb{A}$ is assessing her risk[10] using the entropic risk measure ρ_0^a for which the driver $g^a : \mathbb{R}^2 \to \mathbb{R}$ is given by

$$g^a(z) := \frac{1}{2\gamma_a}|z|^2, \quad \text{where} \quad \gamma_a > 0 \quad \text{is agent } a\text{'s } \textit{risk tolerance}, \tag{2.4.15}$$

[9] Recall argument: If π^* is integrable against prices and solves (2.4.12), then by (2.4.9) we have $\pi_t^{*,a} = \Pi^a(t, \pi_t^{*,-a}, \widetilde{Z}_t^a)$, hence Theorem 2.4.2 tells us that $\pi^{*,a}$ is the best response to $\pi^{*,-a}$, and we therefore have a NE.

[10] Our problem is that of minimizing a static risk measure. The corresponding dynamic risk measure is be given by $\rho_t^a(\xi) := \gamma_a \ln \mathbb{E}[e^{-\xi/\gamma_a}|\mathcal{F}_t]$ for $t \in [0,T]$.

and $1/\gamma_a$ is agent a's *risk aversion* (see [BE09, Proposition 3.12]).

Remark 2.4.8. *The negative of the entropic risk measure is a certainty equivalent for exponential utility (see e.g.* [BE09, Section 3.1.1] *or* [FK11, Page 336]*):*

$$\rho_0^a(\xi) = Y_0^a = \gamma_a \ln \mathbb{E}[e^{-\xi/\gamma_a}] \quad \textit{satisfies} \quad u^a(-\rho_0^a(\xi)) = \mathbb{E}[u^a(\xi)], \qquad \textit{where} \quad u^a(\xi) = -\gamma_a e^{-\xi/\gamma_a}.$$

Thus, minimizing an agent's entropic risk is equivalent to maximizing her expected exponential utility. Furthermore, exponential utility / entropic risk has the advantageous structure of base preference functionals (see [CHKP16, Equation (24)]*) or γ-tolerant risk measures (see* [BE09, Section 3.2.1]*). See also* [REK00] *for a pricing problem in a model with (negative) exponential utility.*

Remark 2.4.9. *In Theorems 2.4.2 and 2.4.5 we made the assumption that $|\widetilde{G}^a(t,0)|^{1/2} \in \mathcal{H}_{BMO}$, which ensures the well-posedness of the minimized-risk BSDEs. In this example, for $a \in \mathbb{A}$, $\widetilde{G}^a(t,0) = -\frac{\gamma_a}{2}|\theta_t|^2$, and furthermore $\theta \in \mathcal{H}_{BMO}$ by assumption. Hence $|\widetilde{G}^a(t,0)|^{1/2} \in \mathcal{H}_{BMO}$ is satisfied.*

Since $g^a_{z^i}(z) = z^i/\gamma_a$, it is easily found that $\mathcal{Z}^a(t,-\theta_t) = (-\gamma_a\theta_t^S, -\gamma_a\theta_t^R) = -\gamma_a\theta_t$ for all $a \in \mathbb{A}$ and $i \in \{1,2\}$ (cf. (2.3.8)). Injecting this in (2.4.10) yields the minimized driver $\widetilde{G}^a$,

$$\widetilde{G}^a(t,z^a) = -\frac{\gamma_a}{2}|\theta_t|^2 - \langle z^a, \theta_t \rangle, \quad t \in [0,T]. \tag{2.4.16}$$

The minimized (individual) risk is then given by $Y_0^a = \widetilde{Y}_0^a$ where $(\widetilde{Y}^a, \widetilde{Z}^a)$ is the solution to the BSDE with terminal condition $-H^a$ and driver $\widetilde{G}^a$, while the optimal strategies $\pi^* = (\pi^{*,a})_{a\in\mathbb{A}}$ are given by (cf. (2.4.9))

$$\pi^{*,a,1} - \widetilde{\lambda}^a \sum_{b\in\mathbb{A}\setminus\{a\}} \pi^{*,b,1} = \frac{\widetilde{Z}^{a,1} + \gamma_a\theta^S}{\sigma^S S} - \frac{\widetilde{Z}^{a,2} + \gamma_a\theta^R}{\kappa^R}\frac{\kappa^S}{\sigma^S S}, \tag{2.4.17}$$

$$\pi^{*,a,2} - \widetilde{\lambda}^a \sum_{b\in\mathbb{A}\setminus\{a\}} \pi^{*,b,2} = \frac{\widetilde{Z}^{a,2} + \gamma_a\theta^R}{\kappa^R}. \tag{2.4.18}$$

The general invertibility of the systems (2.4.17) and (2.4.18) given θ is guaranteed by Theorem 2.4.5.

Remark 2.4.10. *If one imposes* (2.3.12) *with $n = 0$, implying that $\sum_{b\in\mathbb{A}\setminus\{a\}} \pi^{*,b,2} = -\pi^{*,a,2}$, then the linear system* (2.4.18) *for the investment in the derivative simplifies greatly and its solution is explicitly given by*

$$\pi^{*,a,2} = \frac{1}{1+\widetilde{\lambda}^a}\frac{\widetilde{Z}^{a,2} + \gamma_a\theta^R}{\kappa^R} \qquad \textit{for all } a \in \mathbb{A}. \tag{2.4.19}$$

In Section 2.7 we study a model with two agents and computations will be done explicitly for the investment in the stock, including an explicit inversion of System (2.4.17).

The case of multiple agents without relative performance concerns

If one assumes $\lambda^a = 0$ for all $a \in \mathbb{A}$, then recovers the optimal strategies from [HPDR10, Section 4.1.1] as special case, namely

$$\pi^{\lambda=0,a,1} := \frac{\widetilde{Z}^{a,1} + \gamma_a \theta^S}{\sigma^S S} - \frac{\widetilde{Z}^{a,2} + \gamma_a \theta^R}{\kappa^R} \frac{\kappa^S}{\sigma^S S} \quad \text{and} \quad \pi^{\lambda=0,a,2} := \frac{\widetilde{Z}^{a,2} + \gamma_a \theta^R}{\kappa^R}. \tag{2.4.20}$$

and as the driver for the minimized residual risk $\widetilde{Y}^a$ is affine, we have the explicit solution

$$\widetilde{Y}^a_0 = \mathbb{E}^{-\theta}\left[-H^a - \frac{\gamma_a}{2}\int_0^T |\theta_u|^2 \mathrm{d}u\right] = -\mathbb{E}\left[\mathcal{E}_T^{-\theta} \cdot \left(H^a + \frac{\gamma_a}{2}\int_0^T |\theta_u|^2 \mathrm{d}u\right)\right].$$

Observe that in this case the strategy $\pi^{\lambda=0,a}$ followed by a does not depend *directly* on the strategies of the other agents. However, when the price dynamics of the derivative is not fixed but emerges from the equilibrium, later on, the other agents' strategies will appear *indirectly* via θ^R and κ.

Remark 2.4.11. *If one assumes $n = 0$, then the market clearing condition* (2.3.12) *reads $\sum_{a\in\mathbb{A}} \pi^{\lambda=0,a,2} = 0$. With this, the market price of external risk θ^R can be computed from adding* (2.4.20) *over $a \in \mathbb{A}$, giving $\theta^R = -\sum_{a\in\mathbb{A}} \widetilde{Z}^{a,2} / \sum_{a\in\mathbb{A}} \gamma_a$. However, each $\widetilde{Z}^{\cdot,2}$ originates from a BSDE involving θ^R. If one replaces θ^R in each BSDE by $-\sum_{a\in\mathbb{A}} \widetilde{Z}^{a,2} / \sum_{a\in\mathbb{A}} \gamma_a$, then one gets a (coupled) system of N BSDEs, each of which depends on all $\widetilde{Z}^{a,2}$ for $a \in \mathbb{A}$. This idea will be presented in detail in Section 2.5.1.*

The reference case of a single agent that cannot trade in the derivative

It is also instructive, and will be useful later on, to look at the case where this single agent cannot trade in the derivative, and hence faces an incomplete market. We first enforce $\pi^{a,2} = 0$ on (2.4.4), then we optimize over $\pi^{a,1}$ (see Remark 2.4.3). The minimized driver following the calculations is

$$\widetilde{G}^a(t, z) = -\frac{\gamma_a}{2}(\theta_t^S)^2 - z^1\,\theta_t^S + \frac{1}{2\gamma_a}(z^2)^2, \quad t \in [0, T]. \tag{2.4.21}$$

Notice that $\widetilde{G}^a$ is affine in the variable z^1 but retains the quadratic term in z^2. The minimized risk is then given by $Y_0^a = \widetilde{Y}_0^a$ where $(\widetilde{Y}^a, \widetilde{Z}^a)$ is the solution to the BSDE with terminal condition $-H^a$ and the above driver $\widetilde{G}^a$, while the optimal strategy is

$$\pi^{*,a,1} = \frac{\widetilde{Z}^{a,1} + \gamma_a \theta^S}{\sigma^S S} \quad \text{and} \quad \pi^{*,a,2} = 0. \tag{2.4.22}$$

2.4.4 Reduction to zero net supply

In this section we give an auxiliary result, which allows to simplify Condition (2.3.12). We show how the initial holdings $\pi_{0-}^{a,2} = \pi_0^{a,2} \neq 0$ before/at the beginning of the game can be reduced to the case where $\pi_{0-}^{a,2} = \pi_0^{a,2} = 0$. This allows us to apply (2.3.12) with $n = 0$, which will prove crucial in later computations. The reduction to $n = 0$ is

based on the monotonicity of the risk measures and the following lemma, stated from the point of view of one agent $a \in \mathbb{A}$. The result is based on Lemma 3.9 in [HPDR10]. We omit its proof as it does not require notable changes.
To avoid a notational overload, we omit explicit dependencies on π^{-a}.

Lemma 2.4.12. *For a given MPR θ and admissible strategies $\pi^{-a} = (\pi^b)_{b\in\mathbb{A}\setminus\{a\}}$, consider the dynamics of the residual risk BSDE*

$$-\mathrm{d}\widetilde{Y}^a_t(\pi^a) = \widetilde{g}^a\big(t, \pi^a_t, \widetilde{Z}^a_t(\pi^a)\big)\mathrm{d}t - \langle \widetilde{Z}^a_t(\pi^a), \mathrm{d}W_t\rangle \tag{2.4.23}$$

associated with the preferences of agent a using an admissible strategy π^a. Assume further that (2.4.23) has a unique solution for any given $\mathcal{F}_T$-measurable bounded terminal condition $\widetilde{Y}_T$. Let $\nu \in \mathbb{R}$. Then,

- *if $\pi^a := (\pi^{a,1}, \pi^{a,2})$ minimizes the solution $\widetilde{Y}_0(\pi^a)$ to (2.4.23) for a terminal condition $-H^a$, then $\check{\pi}^a := (\pi^{a,1}, \pi^{a,2} - \nu)$ is optimal for the terminal condition $-(H^a + \nu H^D)$;*
- *if $\pi^a := (\pi^{a,1}, \pi^{a,2})$ minimizes the solution $\widetilde{Y}_0(\pi^a)$ for a terminal condition $-(H^a + \nu H^D)$, then $\widehat{\pi}^a := (\pi^{a,1}, \pi^{a,2} + \nu)$ is optimal for the terminal condition $-H^a$.*

This lemma intuitively states that an agent a, owning at time $t = 0$ a portion $\nu^a = \pi^{a,2}_{0^-} = \pi^{a,2}_0$ of units of H^D, can be regarded as being in fact endowed with $\check{H}^a = H^a + \nu^a H^D$. One then looks only at the relative portfolio $\check{\pi}^{a,2} = \pi^{a,2} - \nu^a$, which counts the derivatives bought and sold only from time $t = 0$ onwards: the optimization problem is equivalent. The argument can be extended to all other agents. We note that this reduction is only possible because we do not consider any trading constraints, so that either both strategies $\pi^{a,2}$ and $\check{\pi}^{a,2}$ are admissible or neither one is.

Henceforth we assume that each agent receives at $t = T$ a portion[11] n/N of the derivative H^D. By doing so, the market clearing condition in Definition 2.3.5 transforms into

$$\sum_{a\in\mathbb{A}} \pi^{a,2}_t = 0 \quad \mathbb{P}\otimes Leb - a.e.,$$

and we refer to it as the *zero net supply condition*.

For clarity, we recall that agent $a \in \mathbb{A}$ now assesses her risk by solving the dynamics provided by BSDE (2.3.10) with terminal condition

$$Y^a_T = -\Big(H^a + \frac{n}{N}H^D + V^{a,\theta}_T(\pi^a) - \widetilde{\lambda}^a \sum_{b\in\mathbb{A}\setminus\{a\}} V^{b,\theta}_T(\pi^b)\Big) \tag{2.4.24}$$

(instead of that in (2.3.10)). Moreover, by applying the change of variables (2.4.1) to BSDE (2.3.10) with terminal condition (2.4.24), we obtain

$$\begin{aligned} -\mathrm{d}\widetilde{Y}^a_t &= \widetilde{g}^a(t, \pi^a_t, \pi^{-a}_t, \widetilde{Z}^a_t)\mathrm{d}t - \langle \widetilde{Z}^a_t, \mathrm{d}W_t\rangle, \quad t \in [0, T], \\ \widetilde{Y}^a_T &= -\left(H^a + \frac{n}{N}H^D\right), \end{aligned} \tag{2.4.25}$$

[11] Many possibilities for this reduction to zero net supply exist, including endowing one agent with the total amount n of derivatives H^D or endowing each agent with their initial portions of the derivative ν^a. We make the judicious choice of n/N for simplicity.

where $\widetilde{g}^a$ is given by (2.4.4) and $(\widetilde{Y}^a, \widetilde{Z}^a)$ relates to (Y^a, Z^a) via the change of variables (2.4.1).

It is straightforward to recompile the results of Section 2.4.3 under the *zero net supply* condition. It entails no changes in the strategies or drivers, only the terminal conditions of the involved BSDEs need to be updated from $-H^a$ to $-(H^a + \frac{n}{N}H^D)$ as in (2.4.25).

2.5 The equilibrium market price of external risk

In the previous section we saw how to compute the NE for a given MPR $\theta = (\theta^S, \theta^R)$, without the global constraint on trading (market clearing condition). In this section we solve the equilibrium problem, as posed by Definition 2.3.5, by finding the EMPeR θ^R.

The literature contains many results on equilibria in complete markets that link competitive equilibria to an optimization problem for a representative agent, and this is the approach we use here. The preferences of the representative agent are usually given by a weighted average of the individual agents' preferences with the weights depending on the competitive equilibrium to be supported by the representative agent, see [Neg60]. This dependence results in complex fixed point problems, rendering the analysis and computation of equilibria quite cumbersome. The many results on risk sharing under translation invariant preferences, in particular [BE05], [JST06] and [FK08], suggest that when the preferences are translation invariant, then all the weights are equal. This was an effective strategy in [HPDR10] and it would be so here if, for all $a \in \mathbb{A}$, $\lambda^a = \lambda \in [0, 1)$.

In a market without performance concerns, [HPDR10] shows that the infimal convolution of risk measures gives rise to a suitable risk measure for the representative agent which, for g-conditional risk measures, corresponds to infimal convolution of the drivers. Due to the performance concerns, we use a weighted-dilated infimal convolution, and in Theorem 2.5.4 we show that indeed minimizing the risk of our representative agent is equivalent to finding a competitive equilibrium in our market.

2.5.1 The benchmark case of entropic risk measures

Before presenting the representative agent approach in full generality, we will consider the case of entropic risk measures. In Remark 2.4.11 we already mentioned that in this special case, one can obtain a characterization of the EMPR as a linear combination of solutions to the individual BSDEs (2.4.2) with the minimized driver $\widetilde{G}^a$ given by (2.4.16). Applying the ideas from Remark 2.4.11 to Equation (2.4.19), we see that the market clearing condition requires

$$0 = \sum_{a \in \mathbb{A}} \pi^{*,a,2} = \sum_{a \in \mathbb{A}} \frac{1}{1 + \widetilde{\lambda}^a} \frac{\widetilde{Z}^{a,2} + \gamma_a \theta^R}{\kappa^R} \iff \theta^R = -\frac{\sum_{a \in \mathbb{A}} \frac{\widetilde{Z}^{a,2}}{(1+\widetilde{\lambda}^a)}}{\sum_{a \in \mathbb{A}} \frac{\gamma_a}{(1+\widetilde{\lambda}^a)}} = -\frac{\sum_{a \in \mathbb{A}} w^a \widetilde{Z}^{a,2}}{\gamma_R},$$

if we define $\gamma_R := \sum_{a \in \mathbb{A}} w^a \gamma_a$, with $w^a = 1/(\Lambda(1 + \widetilde{\lambda}^a))$ and $\Lambda = \sum_{a \in \mathbb{A}} 1/(1 + \widetilde{\lambda}^a)$. Notice that here we normalize the family $w = (w^a)_{a \in \mathbb{A}}$ so that $\sum_{a \in \mathbb{A}} w^a = 1$ in order to be consistent with the aggregation in the following section. Any rescaling Λ' of w would give the same θ^R.

Replacing the term θ^R by the above value in the minimized driver given by (2.4.16), we find that the optimal risk processes for each agent solve the BSDEs with drivers given by

$$\widetilde{G}^a(t, \widetilde{Z}_t^{\mathbb{A}}) = -\frac{\gamma_a}{2}\left(\theta_t^S\right)^2 - \widetilde{Z}_t^{a,1}\theta_t^S + \frac{1}{\gamma_R}\,\widetilde{Z}_t^{a,2}\left(\sum_{b\in\mathbb{A}} w^b \widetilde{Z}_t^{b,2}\right) - \frac{\gamma_a}{2\gamma_R^2}\left(\sum_{b\in\mathbb{A}} w^b \widetilde{Z}_t^{b,2}\right)^2 \quad (2.5.1)$$

for $t \in [0,T]$. The BSDEs with these drivers form a system of N coupled BSDEs with quadratic growth, which, in general, are difficult to solve[12]. By taking advantage of the structure of (2.5.1) one can find a simpler BSDE for the process $(\widehat{Y}^w, \widehat{Z}^w) := \sum_{a\in\mathbb{A}} w^a(\widetilde{Y}^a, \widetilde{Z}^a)$. It is easily seen that $\widehat{Y}_T^w = -\sum_{a\in\mathbb{A}} w^a(H^a + nH^D/N) =: -H^w$, cf. (2.5.7). Linearly combining the BSDEs (2.4.2) with drivers expressed as in (2.5.1), we find

$$\begin{aligned} -\mathrm{d}\widehat{Y}_t^w &= \Big[-\frac{\gamma_R}{2}\left(\theta_t^S\right)^2 - \widehat{Z}_t^{w,1}\theta_t^S + \frac{1}{2\gamma_R}\left(\widehat{Z}_t^{w,2}\right)^2\Big]\mathrm{d}t - \langle \widehat{Z}_t^w, \mathrm{d}W_t\rangle, \quad t\in[0,T], \\ \widehat{Y}_T^w &= -H^w. \end{aligned} \quad (2.5.2)$$

Given that H^w and θ^S are bounded, this BSDE falls in the standard class of quadratic growth BSDEs and the existence and uniqueness of $(\widehat{Y}^w, \widehat{Z}^w)$ is easily guaranteed. This allows one to compute θ^R as $-\widehat{Z}^{w,2}/\gamma_R$ and in turn one can finally solve the BSDEs giving the minimized risk processes for each agent, using the driver $\widetilde{G}^a$ as given by (2.4.16).

For more general drivers than that of entropic risk, the condition on the EMPR θ would read

$$\sum_{a\in\mathbb{A}} \frac{\widetilde{Z}_t^{a,2}}{1+\widetilde{\lambda}^a} = \sum_{a\in\mathbb{A}} \frac{\mathcal{Z}_t^{a,2}(t,-\theta_t)}{1+\widetilde{\lambda}^a}.$$

Solving for θ is particularly easy for entropic risk, because in that case $\mathcal{Z}^{a,2}(t,-\theta_t) = -\gamma_a\theta_t$. If a given driver was less well behaved in this regard, or if one wanted to obtain a method applicable to all drivers which satisfy our assumptions, then one could still aim for a single BSDE whose solution will be related to θ^R as above. The method of infimal convolution with the weights for the dilation, $(w^a)_{a\in\mathbb{A}}$, which have been suggested above, will be shown to give the desired BSDE and thus the EMPR θ.

2.5.2 The representative agent approach

Aggregation of risks and aggregation of drivers

Inspired by the above mentioned results and having in mind [Rüs13] (see Remark 2.5.9 below) we deal with the added inter-dependency arising from the fixed-supply condition and the additional unknown θ^R (see Remarks 2.4.11 and 2.4.10) by defining a new risk

[12] Based on the works of [Esp10] and [ET15], the authors of [FdR11] give several counter examples to the existence of solutions to system of fully coupled multi-dimensional quadratic BSDEs; nonetheless, under suitable assumptions positive results do exist, see e.g. [Tev08], [Fre14], [JKL14], [CN15], [KP16], [HT16] and [XŽ16].

measure ρ_0^w. For a set of positive weights $w = (w^a)_{a\in\mathbb{A}}$ satisfying $\sum_{a\in\mathbb{A}} w^a = 1$, we define[13] for any $X \in L^\infty(\mathcal{F}_T)$

$$\rho_0^w(X) := \inf\left\{\sum_{a\in\mathbb{A}} w^a\rho_0^a\big(X^a\big) \;\middle|\; (X^a)\in (L^\infty)^N : \sum_{a\in\mathbb{A}} w^a X^a = X\right\}. \tag{2.5.3}$$

In the case of risk measures induced by BSDEs, [BE05] shows that the measure defined by inf-convolution of risk measures $(\rho_0^a)_{a\in\mathbb{A}}$ is again induced by a BSDE, whose driver is simply the inf-convolution of the BSDE drivers g^a for the risk measures $(\rho_0^a)_{a\in\mathbb{A}}$. For a given set of weights $w = (w^a)_{a\in\mathbb{A}}$, we define the driver g^w as the *weighted-dilated* inf-convolution of the drivers g^a, i.e., for $(t,z)\in[0,T]\times\mathbb{R}^2$,

$$\begin{aligned} g^w(t,z) &:= \square_w\Big((g^a)_{a\in\mathbb{A}}\Big)(t,z) \\ &:= \inf\left\{\sum_{a\in\mathbb{A}} w^a g^a(t,z^a) \;\middle|\; (z^a)\in(\mathbb{R}^2)^N \text{ s.t. } \sum_{a\in\mathbb{A}} w^a z^a = z\right\}, \end{aligned} \tag{2.5.4}$$

where the notation $\square\big((g^a)_{a\in\mathbb{A}}\big)$ would be that commonly used for standard infimal convolution (cf. [BE05, Section 3.2]).

Lemma 2.5.1 (Properties of g^w). *The map $g^w : [0,T]\times\mathbb{R}^2\to\mathbb{R}$ defined by* (2.5.4) *is a deterministic continuous function, strictly convex and continuously differentiable. Moreover, there exists a unique solution of $\nabla_z g^w(t,\mathcal{Z}) = -\vartheta$ in $\mathcal{Z}$.*
For $z^{\mathbb{A}} = (z^a)_{a\in\mathbb{A}}$ such that $\sum_{a\in\mathbb{A}} w^a z^a = z$, one has $g^w(t,z) = \sum_{a\in\mathbb{A}} w^a g^a(t,z^a)$ if and only if there exists $\vartheta\in\mathbb{R}^2$ such that, for all $a\in\mathbb{A}$, $\nabla_z g^a(t,z^a) = -\vartheta$. In that case, the FOC for g^w, $\nabla_z g^w(t,z) = -\vartheta$, holds.

Proof. The weighted inf-convolution transfers the properties of the drivers g^a ($a\in\mathbb{A}$) to g^w, in particular continuity, strict convexity and differentiability. We do not show these as they follow from a simple adaptation of known arguments, cf. [BE05, Section 3.2], [BE09, Sections 3.3 and 3.8] and [HPDR10, Assumption 3.10].
Since the function being minimized ($z^{\mathbb{A}} = (z^a)_{a\in\mathbb{A}} \mapsto \sum_{a\in\mathbb{A}} w^a g^a(z^a)$) is strictly convex and the function defining the constraint ($z^{\mathbb{A}} \mapsto \sum_{a\in\mathbb{A}} w^a z^a$) is also convex, because affine, the minimization defining g^w is equivalent to finding a critical point for the associated Lagrangian, $L(z^{\mathbb{A}},\vartheta) = \sum_{a\in\mathbb{A}} w^a g^a(z^a) + \vartheta(\sum_{a\in\mathbb{A}} w^a z^a - z)$. Therefore, $z^{\mathbb{A}} = (z^a)_{a\in\mathbb{A}}$ satisfying $\sum_{a\in\mathbb{A}} w^a z^a = z$ is a minimizer if and only if there exists $\vartheta\in\mathbb{R}^2$ such that, for all $a\in\mathbb{A}$, $\nabla_z g^a(t,z^a) = -\vartheta$. Then, $\nabla_z g^w(t,z) = -\vartheta$ where ϑ is the Lagrange multiplier associated with z. $\square$

The risk of the random terminal wealth ξ^w, measured through ρ_0^w, is given by $\rho_0^w(\xi^w) := Y_0^w$ where (Y^w, Z^w) is the solution to the BSDE

$$\begin{aligned} -\mathrm{d}Y_t^w &= g^w(t,Z_t^w)\mathrm{d}t - \langle Z_t^w, \mathrm{d}W_t\rangle, \quad t\in[0,T], \\ Y_T^w &= -\xi^w. \end{aligned} \tag{2.5.5}$$

Since the weights $(w^a)_{a\in\mathbb{A}}$ are required to satisfy $\sum_{a\in\mathbb{A}} w^a = 1$, the risk measure ρ_0^w associated to the BSDE with the above driver is a monetary risk measure. Translation

[13] For our purpose, it suffices to define the static risk measure ρ_0^w. Once the BSDE is identified, one obtains of course a corresponding dynamic risk measure $(\rho_t^w)_{t\in[0,T]}$.

invariance and monotonicity follow from the fact that the driver g^w is independent of y. Convexity follows from the convexity of g^w, which in turns follows from that of the drivers g^a ($a \in \mathbb{A}$) by the envelope theorem.

Remark 2.5.2. *Notice that* (2.5.4) *can be rewritten*

$$g^w(t,z) = \inf\left\{\sum_{a\in\mathbb{A}} w^a g^a\big(t, \frac{z^a}{w^a}\big) \;\middle|\; \sum_{a\in\mathbb{A}} z^a = z\right\}.$$

In this way, g^w is seen as the usual w-weighted infimal convolution of the w^a-dilated drivers g^a, in the terminology from [BE09, page 137]. *For more on dilated risk measures, see* [BE09, Proposition 3.4].

Example 2.5.3 (Entropic risk measure). *For entropic risk, i.e. drivers given by $g^a(z^a) = \frac{|z^a|^2}{2\gamma_a}$ for $a \in \mathbb{A}$, one obtains*

$$g^w(z) = \frac{|z|^2}{2\gamma_R}, \quad \textit{with} \quad \gamma_R := \sum_{a\in\mathbb{A}} w^a \gamma_a. \tag{2.5.6}$$

Trading and the risky position of the representative agent

Having defined the aggregated risk measure ρ_0^w and the associated driver g^w, we now introduce a strategy π^w and associated trading gains $V_\cdot(\pi^w) = \int_0^\cdot \pi_t^{w,1}\mathrm{d}S_t + \int_0^\cdot \pi_t^{w,2}\mathrm{d}B_t$ for a representative agent whose preferences are described by g^w. Direct computations from (2.5.3) entail that we assign to the representative agent the terminal position

$$\begin{aligned}
\xi^w := \sum_{a\in\mathbb{A}} w^a \xi^a &= \sum_{a\in\mathbb{A}} w^a \left(H^a + \frac{n}{N} H^D + V_T^a - \widetilde{\lambda}^a \bar{V}_T^{-a} \right) \\
&= \sum_{a\in\mathbb{A}} w^a (H^a + \frac{n}{N} H^D) + \sum_{a\in\mathbb{A}} w^a \Big((1+\widetilde{\lambda}^a) V_T^a - \widetilde{\lambda}^a \sum_{b\in\mathbb{A}} V_T^b \Big) \\
&= \sum_{a\in\mathbb{A}} w^a (H^a + \frac{n}{N} H^D) + \sum_{a\in\mathbb{A}} V_T^a \Big(w^a (1+\widetilde{\lambda}^a) - \sum_{b\in\mathbb{A}} w^b \widetilde{\lambda}^b \Big) \\
&=: H^w + V_T(\pi^w),
\end{aligned}$$

where for $c^a := w^a(1+\widetilde{\lambda}^a) - \sum_{b\in\mathbb{A}} w^b \widetilde{\lambda}^b$,

- $\pi^w = \sum_{a\in\mathbb{A}} c^a \pi^a$ is the representative agent's portfolio,
- $V_T(\pi^w) = \sum_{a\in\mathbb{A}} c^a V_T(\pi^a)$ is the representative agent's wealth process and
- H^w, given by terminal condition

$$H^w := \sum_{a\in\mathbb{A}} w^a (H^a + \frac{n}{N} H^D) = \frac{n}{N} H^D + \sum_{a\in\mathbb{A}} w^a H^a, \tag{2.5.7}$$

  is the representative agent's terminal endowment.

We now choose the weights $(w^a)_{a\in\mathbb{A}}$ such that $c^a = c$ for any $a \in \mathbb{A}$ for some $c \in (0, +\infty)$, namely,

$$w^a := \frac{1}{\Lambda(1+\widetilde{\lambda}^a)} \quad \text{for all } a \in \mathbb{A}, \text{ where} \quad \Lambda := \sum_{a\in\mathbb{A}} \frac{1}{1+\widetilde{\lambda}^a}. \tag{2.5.8}$$

Direct verification yields $\sum_a w^a = 1$ and, furthermore, for all $a \in \mathbb{A}$,

$$c^a = w^a(1+\widetilde{\lambda}^a) - \sum_{b\in\mathbb{A}} w^b\widetilde{\lambda}^b = \frac{1}{\Lambda}\left(1 - \sum_{b\in\mathbb{A}} \frac{\widetilde{\lambda}^b}{1+\widetilde{\lambda}^b}\right) =: c,$$

i.e., c^a does really not depend on a and hence loses rightfully its superindex.

Notice that $\pi^{w,2} = \sum_{a\in\mathbb{A}} c^a\pi^{a,2} = c\sum_{a\in\mathbb{A}} \pi^{a,2}$. In other words, the zero net supply condition for the individual agents (i.e. $\sum_{a\in\mathbb{A}} \pi^{a,2} = 0$) is equivalent to the representative agent not investing in H^D (i.e. $\pi^{w,2} = 0$). This choice is necessary for successful aggregation of risks and simultaneously keeping the zero net supply condition intact. From now on, the family of weights w is fixed and is given by (2.5.8).

The pointwise minimizer for the representative agent's residual risk

We now show that the approach by aggregated risk and representative agent, as motivated above, allows to identify the equilibrium market price of risk as a by-product of minimizing the risk of the representative agent. This risk is given by the solution to BSDE (2.5.5) with terminal condition $Y_T^w = -\xi^w = -H^w - V_T(\pi^w)$, for admissible strategies π^w of the form $\pi^w = (\pi^{w,1}, 0)$. The $\mathbb{R}^2$-valued strategy process π^w is said to be admissible ($\pi^w \in \mathcal{A}^w$) if $\mathbb{E}^\theta\left[\langle\, V_\cdot(\pi^w)\,\rangle_T\right] < \infty$ and if the BSDE (2.5.5) has a unique solution. Following Section 2.4 we introduce the residual risk processes

$$\widetilde{Y}_t^w := Y_t^w + V_t^w \quad \text{and} \quad \widetilde{Z}_t^w := Z_t^w + \pi_t^{w,1}\sigma_t, \quad t \in [0,T].$$

The pair $(\widetilde{Y}^w, \widetilde{Z}^w)$ satisfies the BSDE with terminal condition $\widetilde{Y}_T^w = -H^w$ and random driver $\widetilde{g}^w$, defined for $(\omega, t, \pi_t^w, z) \in \Omega \times [0,T] \times \mathbb{R}^2 \times \mathbb{R}^2$, by

$$\widetilde{g}^w(t, \pi_t^w, z) := g^w\big(t, z - \zeta_t^w\big) - \langle \zeta_t^w\,,\, \theta_t\rangle\,,$$

where $\zeta^w = \pi^{w,1}\sigma$ (compare with (2.4.1)-(2.4.4)). Since $\widetilde{Y}_0^w = Y_0^w$, the representative agent then equivalently aims at solving (i.e., finding the minimizer to)

$$\min_{\pi^w\in\mathcal{A}^w} \widetilde{Y}_0^w(\pi^w).$$

Following the methodology used for the single agent in Section 2.4, we first minimize the driver $\widetilde{g}^w$ pointwise. We define $\Pi^{w,1}(t,z)$ as the optimizer for $\min\big\{\widetilde{g}^w(t,(\pi,0),z)\,|\,\pi \in \mathbb{R}\big\}$, setting $\Pi^{w,2}(t,z) = 0$ as to enforce the zero-net supply condition. Since g^w is strictly convex, so is the function $\widetilde{g}^w$, and the minimum is characterized by the solution to first-order condition

$$g_{z^1}^w\left(t, z - \Pi^{w,1}(t,z)\,\sigma_t\right) = -\theta_t^S, \quad t \in [0,T].$$

We denote the minimized (random) driver by

$$\widetilde{G}^w(t,z) = \widetilde{g}^w\big(t, \Pi^w(t,z), z\big), \quad t \in [0,T]. \tag{2.5.9}$$

Optimality for the representative agent and the equilibrium market price of external risk

We assume that the BSDE with driver $\widetilde{G}^w$ defined in (2.5.9) and terminal condition $-H^w$ has a unique solution $(\widetilde{Y}^w, \widetilde{Z}^w)$ in $\mathcal{S}^\infty \times \mathcal{H}_{\mathrm{BMO}}$. Define the strategy $\pi^{*,w}$ by $\pi^{*,w}_t := \big(\Pi^{w,1}(\omega, t, \widetilde{Z}^w_t), 0\big)$ for $t \in [0,T]$. Like for the individual agents in Section 2.4, the following theorem asserts that $\pi^{*,w}$ is the optimal strategy and $\widetilde{Y}^w_0$ is the minimized risk for the representative agent. Moreover, the theorem relates the EMPR $\theta = (\theta^S, \theta^R)$ (recall Definition 2.3.5) to the solution of the representative agent's optimization problem. Recall the family of weights w given by (2.5.8).

Theorem 2.5.4. *Assume that*

- *the BSDE with driver $\widetilde{G}^w$,* (2.5.9)*, and $\widetilde{Y}^w_T = -H^w$ has a unique solution $(\widetilde{Y}^w, \widetilde{Z}^w)$ in $\mathcal{S}^\infty \times \mathcal{H}_{BMO}$,*
- *the comparison theorem holds for the BSDE with driver $\widetilde{G}^w$,*
- *$\pi^{*,w}_\cdot = \big(\Pi^{w,1}(\cdot, \widetilde{Z}^w_\cdot), 0\big)$ is integrable against the prices S and B,*

then $\widetilde{Y}^w_0$ is the minimal risk for the representative agent and $\pi^{*,w}$ is the unique optimal strategy that leads to this level of risk.
If, for the process $\theta^* = (\theta^S, \theta^R)$, with θ^R defined by

$$g^w_{z^2}\left(t, \widetilde{Z}^w_t - \pi^{*,w,1}_t \sigma_t\right) = -\theta^R_t, \quad t \in [0,T], \tag{2.5.10}$$

the conditions of Theorem 2.4.5 hold, then θ^* is the unique EMPR for the agents in $\mathbb{A}$. Additionally, the minimized aggregated risk $\widetilde{Y}^w$ is linked to the individual minimized risks $(\widetilde{Y}^a)_{a\in\mathbb{A}}$ through the identity $\widetilde{Y}^w = \sum_{a\in\mathbb{A}} w^a \widetilde{Y}^a$. Moreover, the NE for the agents in $\mathbb{A}$ satisfies $\pi^{*,w} = c\sum_{a\in\mathbb{A}} \pi^{*,a}$.

Remark 2.5.5. *The uniqueness of the EMPR θ^* does also hold with respect to the choice of the weights $(w^a)_{a\in\mathbb{A}}$. We chose the weights s.t. $\sum_{a\in\mathbb{A}} w^a = 1$, but this property is not necessary. Thus, the weights are only unique up to multiplication by a (non-zero) constant. To see that a different weights vector would not alter the EMPR, let θ^* be associated to the weights $(w^a)_{a\in\mathbb{A}}$ and consider the new weights $(v^a)_{a\in\mathbb{A}}$ given by $v^a := \varphi w^a$ for $a \in \mathbb{A}$ and a constant $\varphi > 0$.*[14] *If we apply these new weights, then we denote the aggregated variables or processes by carrying superindex v. Thus one can verify that $\pi^v = \varphi\pi^w$, $V^v_T = \varphi V^w_T$, $\zeta^v = \varphi\zeta^w$ etc. In the same manner the drivers change: $g^v(t,z) = \varphi g^w(t, \frac{z}{\varphi})$ and thus $g^v_z(t,z) = g^w_z(t, \frac{z}{\varphi})$. With the new driver and the corresponding $\widetilde{g}^v$ one obtains the new pair $(\widetilde{Y}^v, \widetilde{Z}^v)$ with the property $\widetilde{Z}^v = \varphi\widetilde{Z}^w$. Thus*

$$-\theta^* = g^w_z(t, \widetilde{Z}^w - \zeta^w) = g^v_z(t, \widetilde{Z}^v - \zeta^v),$$

which implies the uniqueness of θ^* with respect to the choice of the weights.

[14] Negative weights would change the interpretation; therefore, only positive weights should be used.

Example 2.5.6 (The entropic case). *In the entropic case, we have found* $g^w(z) = \frac{|z|^2}{2\gamma_R}$, *so we have* $\mathcal{Z}^{w,1}(t, -\theta^S_t) = -\gamma_R \theta^S_t$. *The minimized driver is then*

$$\widetilde{G}^w(t,z) = -\frac{\gamma_R}{2}(\theta^S_t)^2 - z^1\theta^S_t + \frac{1}{2\gamma_R}(z^2)^2,$$

as was found in Subsection 2.4.3, Equation (2.4.21). *This driver is quadratic and regular, and the terminal condition* $-H^w$ *is bounded. From* [Kob00], [IDR10] *there is a unique solution* $(\widetilde{Y}^w, \widetilde{Z}^w)$ *in* $\mathcal{S}^\infty \times \mathcal{H}_{BMO}$ *and the comparison theorem applies (see e.g.* [Kob00, Theorem 2.6], [MY10, Theorem 3.2]*). The optimal strategies are*

$$\pi^{*,w,1} = \frac{\widetilde{Z}^{w,1} + \gamma_R \theta^S}{\sigma^S S} \text{ and } \pi^{*,w,2} = 0.$$

With $\widetilde{Z}^w \in \mathcal{H}_{BMO}$ *and* θ^S *bounded,* $\pi^{*,w,1}$ *is integrable against* S. *This verifies the first three assumptions of the theorem. Furthermore, with* (2.5.10) *and since* $\widetilde{Z}^w \in \mathcal{H}_{BMO}$ *and* θ^S *is bounded, we find that*

$$\theta^R = -\frac{\widetilde{Z}^{w,2}}{\gamma_R} \quad \text{and} \quad \theta^* = (\theta^S, \theta^R) \in \mathcal{H}_{BMO}. \tag{2.5.11}$$

Following on Remark 2.4.3, the optimality of $\pi^{*,w}$ and $(\widetilde{Y}^w, \widetilde{Z}^w)$, for an agent w with preferences described by g^w who trades in S, is obtained exactly in the same way as the optimality for a single agent $a \in \mathbb{A}$ in Theorem 2.4.2. So we prove only the claims of Theorem 2.5.4 related to the EMPR θ^*. First, however, we state a counterpart to Lemma 2.4.4 to the case when no trading in B is possible.

Lemma 2.5.7. *Under the assumptions of Theorem 2.5.4, let* $\widehat{\pi}^w = (\widehat{\pi}^{w,1}, 0)$ *be an admissible strategy and* $(\widehat{Y}^w, \widehat{Z}^w)$ *be the associated risk process, i.e., the solution to the BSDE with driver* $\widetilde{g}^w(t, \widehat{\pi}^w_t, \cdot)$ *and terminal condition* $-H^w$. *Assume that the FOC holds for these processes, i.e., for* $t \in [0,T]$,

$$g^w_{z^1}(t, \widehat{Z}^w_t - \widehat{\zeta}^w_t) = -\theta^S_t \qquad \text{with} \qquad \widehat{\zeta}^w_t := \widehat{\pi}^{w,1}_t \sigma_t.$$

Then $(\widehat{Y}^w, \widehat{Z}^w) = (\widetilde{Y}^w, \widetilde{Z}^w)$ *and* $\widehat{\pi}^w = \pi^{*,w}$.

Proof. Recalling the properties of g^w (see Lemma 2.5.1) and the definition of $\Pi^{w,1}$, the condition $g^w_{z^1}(t, \widehat{Z}^w_t - \widehat{\pi}^{w,1}_t \sigma_t) = -\theta^S_t$ means that $\widehat{\pi}^{w,1}_t = \Pi^{w,1}(t, \widehat{Z}^w_t)$. We have then $\widetilde{g}^w(t, \widehat{\pi}^w_t, \widehat{Z}^w_t) = \widetilde{G}^w(t, \widehat{Z}^w_t)$ (recall (2.5.9)). By the assumed uniqueness of the solution to the BSDE with driver $\widetilde{G}^w(t, \cdot)$ and terminal condition $-H^w$, we have $(\widehat{Y}^w, \widehat{Z}^w) = (\widetilde{Y}^w, \widetilde{Z}^w)$. Consequently, by the uniqueness of the FOC's solution, $\widehat{\pi}^{w,1}_t = \Pi^{w,1}(t, \widehat{Z}^w_t) = \Pi^{w,1}(t, \widetilde{Z}^w_t) = \pi^{*,w,1}_t$. Since both strategies have second component equal to zero, we have $\widehat{\pi}^w = \pi^{*,w}$. □

The next result, to be used in the proof of Theorem 2.5.4, states that aggregating the solutions to the individual optimization problems leads to an optimum for the aggregated preference g^w and identifies the BSDE of the aggregation with the weighted sum of the agents' BSDEs.

Lemma 2.5.8. *Let $\vartheta \in \mathcal{H}_{BMO}$ be a MPR and assume the conditions of Theorem 2.4.5. Let $\pi^* := (\pi^{*,a})_{a\in\mathbb{A}}$ be the unconstrained NE associated with ϑ, and let $(\widetilde{Y}^a, \widetilde{Z}^a)$ be the solution to the minimized-risk BSDE for each agent $a \in \mathbb{A}$, i.e., the terminal condition is $-H^a$ and the driver $\widetilde{G}^a$ is given by* (2.4.10). *Define $(\widehat{Y}^w, \widehat{Z}^w) := \sum_{a\in\mathbb{A}} w^a(\widetilde{Y}^a, \widetilde{Z}^a)$ and $\widehat{\pi}^w := \sum_{a\in\mathbb{A}} c^a\pi^{*,a} = c\sum_{a\in\mathbb{A}} \pi^{*,a}$.*
Then $(\widehat{Y}^w, \widehat{Z}^w)$ and $\widehat{\pi}^w$ are the minimal risk and optimal strategy, respectively, for a single agent whose preferences are given by g^w, who can invest in (S, B) (without trading constraints).

Proof. Firstly, we sum the individual risk BSDEs to obtain $(\widehat{Y}^w, \widehat{Z}^w)$ and its BSDE. We have $\widehat{Y}^w_T = -\sum_{a\in\mathbb{A}} w^a H^a = -H^w$ and

$$\begin{aligned}
-\mathrm{d}\widehat{Y}^w_t &= \Big[\sum_{a\in\mathbb{A}} w^a \widetilde{G}^a(t, \widetilde{Z}^a_t)\Big]\mathrm{d}t - \sum_{a\in\mathbb{A}} w^a\langle \widetilde{Z}^a_t, \mathrm{d}W_t\rangle \\
&= \Big[\sum_{a\in\mathbb{A}} w^a\Big\{g^a\Big(t, \widetilde{Z}^a_t - \zeta^a_t(\pi^*)\Big) - \langle\zeta^a_t(\pi^*), \vartheta_t\rangle\Big\}\Big]\mathrm{d}t - \langle\sum_{a\in\mathbb{A}} w^a\widetilde{Z}^a_t, \mathrm{d}W_t\rangle \\
&= \Big[\sum_{a\in\mathbb{A}} w^a g^a\Big(t, \widetilde{Z}^a_t - \zeta^a_t(\pi^*)\Big) - \Big\langle\widehat{\zeta}^w_t, \vartheta_t\Big\rangle\Big]\mathrm{d}t - \langle\widehat{Z}^w_t, \mathrm{d}W_t\rangle,
\end{aligned}$$

where $\widehat{\zeta}^w := \sum_{a\in\mathbb{A}} w^a\zeta^a(\pi^*) = \widehat{\pi}^{w,1}\sigma + \widehat{\pi}^{w,2}\kappa$. We remark that, on the one hand,

$$\sum_{a\in\mathbb{A}} w^a(\widetilde{Z}^a_t - \zeta^a_t(\pi^*)) = \widehat{Z}^w_t - \widehat{\zeta}^w_t, \quad t \in [0, T],$$

and, on the other hand, from $\pi^{*,a}_\cdot = \Pi^a(\cdot, \pi^{*,-a}_\cdot, \widetilde{Z}^a_\cdot)$ (because π^* is a NE) we can infer that, for all $a \in \mathbb{A}$,

$$\nabla_z g^a\big(t, \widetilde{Z}^a_t - \zeta^a_t(\pi^*)\big) = -\vartheta_t, \quad t \in [0, T].$$

Therefore, we know by Lemma 2.5.1 that $g^w(t, \widehat{Z}^w_t - \widehat{\zeta}^w_t) = \sum_{a\in\mathbb{A}} w^a g^a(t, \widetilde{Z}^a_t - \zeta^a_t(\pi^*))$. This implies that for $t \in [0, T]$

$$\begin{aligned}
-\mathrm{d}\widehat{Y}^w_t &= \big[g^w(t, \widehat{Z}^w_t - \widehat{\zeta}^w_t) - \langle\widehat{\zeta}^w_t, \vartheta_t\rangle\big]\mathrm{d}t - \langle\widehat{Z}^w_t, \mathrm{d}W_t\rangle \\
&= \widetilde{g}^w(t, \widehat{\pi}^w_t, \widehat{Z}^w_t)\mathrm{d}t - \langle\widehat{Z}^w_t, \mathrm{d}W_t\rangle.
\end{aligned}$$

Secondly, by Lemma 2.5.1, we also know that $\nabla_z g^w\big(t, \widehat{Z}^w_t - \widehat{\zeta}^w_t\big) = -\vartheta_t$. Therefore, by Lemma 2.4.4, we obtain that $(\widehat{Y}^w, \widehat{Z}^w)$ is the solution to the minimized-risk BSDE for an agent with preferences given by g^w, terminal condition $-H^w$, who trades in S and B under the given MPR ϑ with $\widehat{\pi}^w$ as the optimal strategy. □

Proof of Theorem 2.5.4. The first part of the proof of the theorem, the optimization for the representative agent, follows through arguments similar to those used in the single agent case, see Theorem 2.4.2 and Remark 2.4.3. Hence we omit it.

▷ *Existence of the EMPR.* Here we prove that $\theta^* = (\theta^S, \theta^R)$, defined through (2.5.10), is indeed an EMPR. Since $\theta^* \in \mathcal{H}_{\text{BMO}}$ and the conditions of Theorem 2.4.5 hold, let $(\pi^{*,a})_{a\in\mathbb{A}}$ be the unique unconstrained NE under the MPR θ^*, and let $(\widetilde{Y}^a, \widetilde{Z}^a)$ be the solution to the minimized-risk BSDE for each agent $a \in \mathbb{A}$. Our goal is to prove that $\sum_{a\in\mathbb{A}} \pi^{*,a,2} = 0$.

Let us introduce $(\widehat{Y}^w, \widehat{Z}^w) := \sum_{a\in\mathbb{A}} w^a(\widetilde{Y}^a, \widetilde{Z}^a)$ and $\widehat{\pi}^w := \sum_{a\in\mathbb{A}} c^a \pi^{*,a} = c\sum_{a\in\mathbb{A}} \pi^{*,a}$. From Lemma 2.5.8, we know that $\widehat{\pi}^w$ and $\widehat{Y}^w$ are the optimal strategy and risk for a single agent with risk preferences encoded by g^w, exposed to terminal risk $-H^w$, trading S and B under θ^* without trading constraints.

Meanwhile, we defined $\pi^{*,w} = (\pi^{*,w,1}, 0)$ as the optimal strategy for an agent w with preferences encoded by g^w and who can only invest in S (with MPR θ^S). By construction of θ^R, we have

$$\nabla_z g^w\big(t, \widetilde{Z}^w_t - \zeta^w_t\big) = -\theta^*_t, \quad \text{with} \quad \zeta^w_t = \pi^{*,w,1}_t \sigma_t, \quad t \in [0,T].$$

From Lemma 2.4.4 we infer that $\pi^{*,w}$ is also the optimal strategy for an agent with preferences g^w and who can invest in S and B with given MPR θ^*. By the uniqueness in Lemma 2.4.4 we therefore have $\widehat{\pi}^w = \pi^{*,w}$. This implies in particular that $\sum_{a\in\mathbb{A}} \pi^{*,a,2} = \widehat{\pi}^{w,2} = \pi^{*,w,2} = 0$. We have thus proved that the NE associated with θ^* satisfies the zero-net supply condition, hence the constructed θ^* is an EMPR.

▷ *Uniqueness of the EMPR.* Assume that $\vartheta := (\theta^S, \vartheta^R)$ is also an EMPR and let $(\pi^{*,a,\vartheta})_{a\in\mathbb{A}}$ be the associated NE for which, by definition of EMPR, the zero-net supply condition $\sum_{a\in\mathbb{A}} \pi^{*,a,\vartheta,2} = 0$ is satisfied. Let also $(\widetilde{Y}^{a,\vartheta}, \widetilde{Z}^{a,\vartheta})$ be the solution to the minimized-risk BSDE for each agent $a \in \mathbb{A}$. As above, we define $(\widehat{Y}^{w,\vartheta}, \widehat{Z}^{w,\vartheta}) := \sum_{a\in\mathbb{A}} w^a(\widetilde{Y}^{a,\vartheta}, \widetilde{Z}^{a,\vartheta})$ and $\widehat{\pi}^{w,\vartheta} := \sum_{a\in\mathbb{A}} c^a \pi^{*,a,\vartheta} = c\sum_{a\in\mathbb{A}} \pi^{*,a,\vartheta}$. By Lemma 2.5.8, we obtain that $(\widehat{Y}^{w,\vartheta}, \widehat{Z}^{w,\vartheta})$ and $\widehat{\pi}^{w,\vartheta}$ are optimal for an agent w who trades in S and B under the given MPR ϑ for a single agent economy. Consequently, using the characterization between the optimizer and the FOC condition, we have

$$g^w_{z^1}\big(t, \widehat{Z}^{w,\vartheta}_t - \widehat{\pi}^{w,\vartheta,1}_t \sigma_t\big) = -\theta^S_t \qquad \text{and} \qquad g^w_{z^2}\big(t, \widehat{Z}^{w,\vartheta}_t - \widehat{\pi}^{w,\vartheta,1}_t \sigma_t\big) = -\vartheta^R_t,$$

where $\widehat{\pi}^{w,\vartheta,2} = 0$ as ϑ is an EMPR. By Lemma 2.5.7, the first equation guarantees that $(\widehat{Y}^{w,\vartheta}, \widehat{Z}^{w,\vartheta})$ and $\widehat{\pi}^{w,\vartheta}$ are optimal for an agent with preferences g^w who trades in S. By the construction of $(\widetilde{Y}^w, \widetilde{Z}^w)$ and $\pi^{*,w}$ (for the MPR θ^*), and the uniqueness recalled in Lemma 2.5.7, we have $(\widehat{Y}^{w,\vartheta}, \widehat{Z}^{w,\vartheta}) = (\widetilde{Y}^w, \widetilde{Z}^w)$ and $\widehat{\pi}^{w,\vartheta} = \pi^{*,w}$. As a consequence, we have from the second FOC equation

$$-\vartheta^R_t = g^w_{z^2}\big(t, \widehat{Z}^{w,\vartheta}_t - \widehat{\pi}^{w,\vartheta,1}_t \sigma_t\big) = g^w_{z^2}\big(t, \widetilde{Z}^w_t - \pi^{*,w,1}_t \sigma_t\big) = -\theta^R_t.$$

Hence the uniqueness of the EMPR θ^*. □

From Theorem 2.5.4 we point out that θ^* is only *a* MPR for the representative agent's economy as the representative agent trades in an incomplete market where she is not able trade the risk from $(R_t)_{t\in[0,T]}$ — recall (2.3.1). Nonetheless, θ^* is the only MPR leading to a complete market such that the NE from the agents' strategies satisfies the zero net supply condition.

In order to put the above result into perspective, we remark that in a complete Arrow-Debreu model of price-taking agents, the representative agent approach is known to give Walrasian equilibrium prices, which motivated this approach in [HPDR10]. In our setting with strategic interaction, however, the success of this method was not quite as predictable.

Remark 2.5.9. *In* [Rüs13, Page 271, Equation (11.25)] *a "weighted minimal convolution" of risk measures is introduced via*

$$\left(\bigwedge \rho_i\right)_\gamma (X) := \inf \left\{ \sum_{i=1}^{N} \gamma_i \rho_i(X_i) \,\middle|\, X_1, \ldots, X_N \in L^p,\ \sum_{i=1}^{N} X_i = X \right\}$$

for $\gamma = (\gamma_i) \in \mathbb{R}^N_{>0}$ and for some $p \geq 1$.
Observe that aggregation in our context would not work without the dilation weights $1/w^a$ in the argument of the driver. This can be seen in the proof of Lemma 2.5.8. The reason is that $\widetilde{G}^a$ is the sum of g^a with the strategies plugged in as arguments and of an additional term containing the strategies multiplied by the weights.

Remark 2.5.10 (No trade-off between risk tolerance and performance concern rate)**.** *Each agent's individual preferences are specified by the parameters γ_a and λ^a, i.e. risk tolerance and performance concern, respectively. One may ask whether a parametric relation between those parameters exists such that an agent with (γ_a, λ^a) and another agent with (γ_b, λ^b) would exhibit the same behavior and have the same optimal strategies. Indeed, in most formulas the two parameters appear as coupled. However, one can see that the terminal condition H^w is independent of the risk tolerance parameter $\gamma_\cdot$, hence by changing λ^a and γ_a of any one fixed agent $a \in \mathbb{A}$, one cannot obtain the same outcome.*

2.6 Further results on the entropic risk measure case

In this section we investigate further the case of entropic risk, i.e., each agent's individual driver is given by (2.4.15). The advantage of this choice is, beside providing a simple set of parametrized drivers, that we can refer to [HPDR10] for the analysis of the model without performance concerns. We refer in particular to Section 4 in [HPDR10]. We introduce a structure that allows to use the theory developed in the previous section and, moreover, to design H^D such that Assumption 2.4.1 holds true. The ultimate goal of this section is to understand how the concern rates $\lambda^\cdot$ affect prices and risks. The first two parts of the section verify that Assumption 2.4.1 holds and the third sheds light on the behavior of the aggregated risk and derivative price as the parameters vary.

We now make further assumptions (commented below) on the structure of the random variables introduced Section 2.3. Namely, we assume that the endowments H^a for $a \in \mathbb{A}$ and the derivative H^D have the form

$$H^a = h^a(S_T, R_T) \quad \text{and} \quad H^D = h^D(S_T, R_T), \tag{2.6.1}$$

respectively, for some deterministic functions $h^\cdot$. This structure for the derivative and endowments is interpreted as each agent receiving a lump sum at maturity time T. To ease the analysis we will assume throughout a Black-Scholes market (i.e., μ^S, σ^S are constants). Throughout the rest of this section the following assumption holds.

Assumption 2.6.1. *Let Assumption 2.3.1 hold. Let $\sigma^S \in (0, \infty)$ and $\mu^R, \mu^S \in \mathbb{R}$ (and hence also $\theta^S \in \mathbb{R}$). For any $a \in \mathbb{A}$ the functions $h^D, h^a \in C^1_b(\mathbb{R}^2; \mathbb{R})$ are strictly positive, their derivatives are uniformly Lipschitz continuous w.r.t. the non-financial risk and satisfy $(\partial_{x_2} h^D)(x_1, x_2) \neq 0$ for any $(x_1, x_2) \in \mathbb{R} \times \mathbb{R}$.*

Remark 2.6.2. *In particular, as $h^D \in C_b^1(\mathbb{R}^2;\mathbb{R})$, $\partial_{x_2} h^D$ is continuous in both arguments. Thus, if it was positive at one and negative at another point, by the intermediate value theorem, it would have to be zero somewhere in between these points, which would contradict the assumption. Therefore, $\partial_{x_2} h^D$ must be either always positive or always negative.*

The assumption concerning the strict positivity of the involved maps or that $\partial_{x_2} h^D \neq 0$ are the key in proving that Assumption 2.4.1 is indeed verified. The assumption on the form of H^a and H^D reduces the BSDE to the Markovian case, giving us access to the many existing BSDE regularity results, which we will use below in their full scope. It would be possible (this is left open to future research) to remain in the non-Markovian setting of general $\mathcal{F}_T$-measurable H^D and H^a and use the link between non-Markovian BSDEs and path-dependent PDEs (see e.g. [EKTZ14]). Indeed, tools on general Malliavin differentiability of BSDE solutions in the non-Markovian setting can be found in [AIdR10] or in more generality in [dR11] and [MPR17].

We recall that our goal is to analyze the impact of the parameters $\lambda^{\cdot}, n, \gamma_{\cdot}$ on the risk processes (single and representative agent), derivative price process and EMPeR.
In this section we work mainly with the representative agent BSDE (see Example 2.5.6 or Equation (2.5.2)) and the derivative price BSDE (2.3.6).

To avoid a notation overload when working with the BSDE for the representative agent, we drop the tilde notation and define (Y^w, Z^w) as the solution to the mentioned BSDE, not to be confused with (2.5.5) which plays no role here. The solution to the derivative price BSDE (2.3.6) is denoted by (B, κ).

2.6.1 The aggregated risk

The BSDE (2.5.2) is not difficult to analyze given the existing literature on BSDEs of quadratic growth. Recall that $\theta^S \in \mathcal{S}^\infty$ and $Y_T^w \in L^\infty$ (since it is a weighted sum of bounded random variables). We shortly recall that $\mathbb{D}^{1,2}$ is the space of first order Malliavin differentiable processes and D denotes the Malliavin derivative operator, we point the reader to Appendix A.2 for further Malliavin calculus references.

Theorem 2.6.3. *The BSDE* (2.5.2) *has a unique solution $(Y^w, Z^w) \in (\mathcal{S}^\infty \cap \mathbb{D}^{1,2}) \times (\mathcal{H}_{BMO} \cap \mathbb{D}^{1,2})$. Moreover, there exists a strictly negative function $u^w \in C^{0,1}([0,T] \times \mathbb{R}^2;\mathbb{R})$ such that for any $t \in [0,T]$*

$$Y_t^w = u^w(t, S_t, R_t) \qquad \text{and} \qquad Z_t^{w,2} = (\partial_{x_2} u^w)(t, S_t, R_t) b, \qquad \mathbb{P}\text{-a.s.}.$$

i) For any $r, u \in [0,t]$, $t \in [0,T]$ it holds that

$$D_u^{W^R} Y_t^w = D_r^{W^R} Y_t^w \quad \mathbb{P}\text{-a.s.} \qquad \text{and} \qquad D_u^{W^R} Z_t^w = D_r^{W^R} Z_t^w \quad \mathbb{P} \otimes Leb\text{-a.e.}$$

and in particular $D_t^{W^R} Y_t = Z_t^w$ $\mathbb{P}$-a.s. for any $t \in [0,T]$.

ii) There exists a constant $C > 0$ such that $|Z_t^{w,2}| \leq C$ for any $t \in [0,T]$, i.e. $Z^{w,2} \in \mathcal{S}^\infty$ and $\partial_{x_2} u^w \in C_b([0,T] \times \mathbb{R}^2;\mathbb{R})$. Moreover, $\theta^R \in \mathcal{S}^\infty$.

iii) The process $D_{\cdot}^{W^R} Z^w$ belongs to $\mathcal{H}_{BMO}$.

Proof. Let $a \in \mathbb{A}$ and $0 \le u \le t \le T$. Existence and uniqueness of the stochastic differential equations (SDEs) (2.3.1) and (2.3.2) follow from Proposition A.3.1.
By assumption we have $Y_T^w \in L^\infty$ and $\theta^S \in \mathcal{S}^\infty$ which allows to quote Theorem 2.6 in [IDR10], implying $(Y^w, Z^w) \in \mathcal{S}^\infty \times \mathcal{H}_{\text{BMO}}$. Moreover, given that $Y_T^w < 0$, a strict comparison principle[15] for quadratic BSDEs (e.g. [MY10, Theorem 3.2][16]) yields $Y_t^w < 0$ for any $t \in [0,T]$ and hence $u^w < 0$.
Proposition A.3.1 ensures that the payoffs H^D and H^a, and hence H^w, are Malliavin differentiable with bounded Malliavin derivatives. Combining this further with $\theta^S \in \mathbb{R}$, the Malliavin differentiability of (2.5.2) follows from Theorem 2.9 in [IDR10]. Under Assumption 2.6.1 the results in [IDR10] (or [dR11, Chapter 4]) along with [AIdR10, Theorem 7.6] yield the Markov property for Y^w and the parametric differentiability result for the (quadratic) BSDE.

▷ *Proof of i):* Since $u^w \in C^{0,1}$, by direct application of the Malliavin differential (or by directly applying (A.3.2)) we have for $0 \le u \le t \le T$ (using $D_u^{W^R} R_t = b$)

$$\begin{aligned} D_u^{W^R} Y_t^w &= D_u^{W^R}\big(u^w(t, S_t, R_t)\big) \\ &= (\partial_{x_2} u^w)(t, S_t, R_t)(D_u^{W^R} R_t) = (\partial_{x_2} u^w)(t, S_t, R_t) b = D_t^{W^R} Y_t^w. \end{aligned}$$

It now follows that

$$D_t^{W^R} Y_t^w = D_u^{W^R} Y_t^w = Z_t^w \text{ for any } 0 \le u \le t \le T \ \mathbb{P}\text{-a.s.} \,. \tag{2.6.2}$$

▷ *Proof of ii):* Define now the probability measure $\mathbb{Q}$ (equivalent to $\mathbb{P}$) via

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \mathcal{E}\left(- \int_0^T \left\langle \left(\theta_s^S, -\frac{Z_s^{w,2}}{\gamma_R} \right), \mathrm{d}W_s \right\rangle \right). \tag{2.6.3}$$

With this choice, $W^{\mathbb{Q}} := W + \int_0^\cdot \left(\theta_s^S, -\frac{Z_s^{w,2}}{\gamma_R} \right) \mathrm{d}s$ is a $\mathbb{Q}$-Brownian Motion. The measure $\mathbb{Q}$ is well defined since $\theta^S \in \mathcal{S}^\infty$ and $Z^{w,2} \in \mathcal{H}_{\text{BMO}}$ (see Lemma A.1.1). For $0 \le u \le t \le T$ we have ([AIdR10, Theorem 8.4] or [IDR10, Theorem 2.9])

$$D_u^{W^R} Y_t^w = D_u^{W^R} Y_T^w + \int_t^T \left[-\theta_s^S D_u^{W^R} Z_s^{w,1} + \frac{1}{\gamma_R} Z_s^{w,2} D_u^{W^R} Z_s^{w,2} \right] \mathrm{d}s - \int_t^T \langle D_u^{W^R} Z_s^w, \mathrm{d}W_s \rangle, \tag{2.6.4}$$

$$\text{which implies} \quad D_u^{W^R} Y_t^w = \mathbb{E}^{\mathbb{Q}}\left[D_u^{W^R} Y_T^w \,\middle|\, \mathcal{F}_t \right].$$

The results in Proposition A.3.1 and the definition of Y_T^w imply that $|D_u^{W^R} Y_t^w| < C$. Path regularity results for BSDEs along with their usual representation formulas (see

[15]The (weak) comparison principle, applied to terminal conditions $-H^w < 0$ and 0, allows only to infer that $Y_t^w \le 0$. If, however we had $Y_s^w = 0$ for any time $s < T$, then the strict version of the theorem implies that we would have $Y_T^w = 0$, hence we have a contradiction, proving that indeed $Y_t^w < 0$ for all times $t \in [0,T]$.

[16][KTPZ15, Theorem 6.1] goes further than necessary for our purpose by providing a strict comparison principle for BSDEs with jumps, whereas [Kob00] is not quite sufficient, because the comparison principle therein (Theorem 2.6) is not strict in the sense of [CEP10]. Theorems 3 and 5 in [CEP10] are strict, but the setting of [MY10] is closer to ours and thus easier to adapt.

[IDR10]) yield that $(D^{W^R}_t Y_t) = (Z^2_t) \in \mathcal{S}^\infty$; the boundedness of $\partial_{x_2} u^w$ follows easily[17]. As a consequence, since $Z^{w,2} \in \mathcal{S}^\infty$ and (2.5.11) holds, we also have $\theta^R \in \mathcal{S}^\infty$.

▷ *Proof of iii):* Using now the fact that $\theta^S, Z^{w,2} \in \mathcal{S}^\infty$, we apply Theorem 2.6 in [IDR10] to (2.6.4) and obtain that $D^{W^R}_\cdot Z^w \in \mathcal{H}_{\text{BMO}}$. The BMO norm of $D^{W^R} Z^w$ depends only on some real constants and T, γ_R, $\sup_u \|D^{W^R}_u Y^w_T\|_{L^\infty}$ and $\|(\theta^S, Z^{w,2})\|_{\mathcal{S}^\infty \times \mathcal{S}^\infty}$ (see again [IDR10, Theorem 2.6]). □

In the next result we show that the mapping $x_2 \mapsto (\partial_{x_2} u^w)(t, x_1, x_2)$ is Lipschitz continuous.

Proposition 2.6.4. *For any $(t, x_1) \in [0, T] \times \mathbb{R}$, the map $\mathbb{R} \ni x_2 \mapsto (\partial_{x_2} u^w)(t, x_1, x_2)$ is Lipschitz continuous uniformly in t and x_1. In particular, the process $D^{W^R} Z^w$ is $\mathbb{P}$-a.s. bounded.*

Proof. Denote by R and $\widetilde{R}$ the solutions to (2.3.1) with $R_0 = r_0$ and $\widetilde{R}_0 = \widetilde{r}_0$ respectively; denote as well by (Y^w, Z^w) and $(\widetilde{Y}^w, \widetilde{Z}^w)$ the solutions to BSDE (2.5.2) for the underlying processes R and $\widetilde{R}$, respectively.
Let $0 \le u \le t \le T$ and define $\delta DY := D^{W^R} Y^w - D^{W^R} \widetilde{Y}^w$, $\delta DZ^i := D^{W^R} Z^{w,i} - D^{W^R} \widetilde{Z}^{w,i}$ for $i \in \{1, 2\}$ and $\delta DZ := (\delta DZ^1, \delta DZ^2)$. Then, following from (2.6.4) written under $\mathbb{Q}$ from (2.6.3), we have

$$\delta D_u Y_t = \delta D_u Y_T - \int_t^T \langle \delta D_u Z_s, \mathrm{d}W^{\mathbb{Q}}_s \rangle + \int_t^T \frac{1}{\gamma_R} (Z^{w,2}_s - \widetilde{Z}^{w,2}_s) D^{W^R}_u Z^{w,2}_s \mathrm{d}s.$$

Define now the process[18]

$$e_t := \exp\left(\int_0^t \frac{1}{\gamma_R} D^{W^R}_u Z^{w,2}_s \mathrm{d}s \right), \quad t \in [0, T] \quad \text{with} \quad (e_t)_{t \in [0,T]} \in \mathcal{H}^p, \ \ \forall p > 1, \qquad (2.6.5)$$

where the $\mathcal{H}^p$-integrability of $(e_t)_{t \in [0,T]}$ follows from Lemma A.1.1. Observe next that by the results of Theorem 2.6.3 one has $\delta D_u Y_t = \delta D_t Y_t = Z^{w,2}_t - \widetilde{Z}^{w,2}_t$. Applying Itô's formula to $(e_t \delta D_\cdot Y_t)$, using the just mentioned identity and taking $\mathbb{Q}$-conditional expectations it follows at $u = t = 0$ that

$$\begin{aligned} |(\partial_{x_2} u^w)(0, s_0, r_0) - (\partial_{x_2} u^w)(0, s_0, \widetilde{r}_0)| &= \frac{1}{b} \left| (Z^{w,2}_0 - \widetilde{Z}^{w,2}_0) \right| = \left| \frac{1}{b} \mathbb{E}^{\mathbb{Q}} \left[e_T \delta D_0 Y_T \right] \right| \\ &\le C |r_0 - \widetilde{r}_0|. \end{aligned}$$

This is a consequence of Proposition A.3.1 combined with the fact that $\mathbb{E}^{\mathbb{Q}}[e^p_T]$ $(\forall p > 1)$ is finite due to the BMO properties of $D^{W^R} Z^{w,2}$, see Lemma A.1.1. The constant C is independent of $u, r_0, \widetilde{r}_0$ and s_0.[19] Although $D^{W^R} Z^{w,2}$ is a BMO martingale under $\mathbb{P}$, the integrability still carries under $\mathbb{Q}$; this is the same argument as in the final step of the proof of Lemma 3.1 in [IDR10] (see also Lemma 2.2 and Remark 2.7 of the cited work).

[17]Recall that $\partial_{x_2} u^w(t, S_t, R_t) b = D^{W^R}_t Y^w_t$, which we just proved to be bounded.

[18]We do not add another index u to e_t, because once we take the expectation, the dependence on u vanishes by Theorem 2.6.3 i). Any $u \in [0, s]$ would give the same expectation.

[19]u appears only indirectly via e_T and the expectation of the latter is independent of u by Theorem 2.6.3 i).

The extension of the above result to the whole time interval $[0,T]$ follows via the Markov property of the BSDE solution. This relates to the close link between BSDEs of the Markovian type and certain classes of quasi-linear parabolic PDEs (see e.g. [EPQ97, Section 4]).

Finally, the boundedness of $D^{W^R}Z^w$ follows from the Lipschitz property of the map $x_2 \mapsto (\partial_{x_2}u^w)(\cdot,\cdot,x_2)$ and the boundedness of $D^{W^R}R$, see Proposition A.3.1, ii). $\square$

2.6.2 The EMPR and the derivative's BSDE

The next result shows that Assumption 2.6.1 on H^a and H^D implies that Assumption 2.4.1 holds for the model with entropic risk, i.e., that the market is complete.

Theorem 2.6.5 (Market completion). *Under the standing assumptions for this chapter, the derivative H^D completes the market, i.e., $\kappa^R \neq 0$ $\mathbb{P}$-a.s. for any $t \in [0,T]$. More precisely, $\kappa^R \in \mathcal{S}^\infty$ and $sgn(\kappa^R_t) = sgn(b\partial_{x_2}h^D)$ for any $t \in [0,T]$.*

Before proving the above result we need an intermediary one. Recall that BSDE (2.3.6) describes the dynamics of the price process B^θ, that $H^D \in L^\infty$ and $\theta \in \mathcal{S}^\infty \times (\mathcal{H}_{\text{BMO}} \cap \mathbb{D}^{1,2})$ (following from Assumption 2.6.1 and Theorem 2.6.3).

Proposition 2.6.6. *The pair (B,κ) belongs to $(\mathcal{S}^\infty \cap \mathbb{D}^{1,2}) \times (\mathcal{H}_{BMO} \cap \mathbb{D}^{1,2})$ and the Malliavin derivatives satisfy for $0 \le u \le t \le T$ the dynamics*

$$D^{W^R}_u B^\theta_t = D^{W^R}_u H^D - \int_t^T \kappa^R_s D^{W^R}_u \theta^R_s + \left\langle \theta_s, D^{W^R}_u \kappa^\theta_s \right\rangle \mathrm{d}s - \int_t^T \langle D^{W^R}_u \kappa^\theta_s, \mathrm{d}W_s \rangle. \quad (2.6.6)$$

The representation $D^{W^R}_t B^\theta_t = \kappa^R_t$ holds $\mathbb{P}$-a.s. for any $0 \le t \le T$.

Proof. Let $0 \le u \le t \le T$. Observe that (2.3.6) (describing the dynamics of B^θ) is a BSDE with a linear driver and a bounded terminal condition. The existence and uniqueness of a solution follows from the results of [EPQ97]. Moreover, the estimation techniques used in [IDR10] yield that $(B,\kappa) \in \mathcal{S}^\infty \times \mathcal{H}_{\text{BMO}}$ (see [IDR10]Theorem 2.6). The Malliavin differentiability of (B,κ) follows from Proposition 5.3 in [EPQ97] and the remark following it since $(\theta^S,\theta^R) \in \mathbb{R} \times (\mathcal{S}^\infty \cap \mathbb{D}^{1,2})$ (see Theorem 2.6.3). [EPQ97, Proposition 5.3] and Proposition A.3.1 yield Equation (2.6.6) and therefore the dynamics of the Malliavin derivative w.r.t. W^R of B^θ. According to [EPQ97, Proposition 5.3], $D^{W^R}_t B^\theta_t$ is a version of κ^θ_t for all $0 \le t \le T$.
We now prove a finer result on B and κ, namely that $D^{W^R}_t B^\theta_t = \kappa^R_t$ holds $\mathbb{P}$-a.s. for any $0 \le t \le T$ instead of just $\mathbb{P} \otimes Leb$-a.e..[20] This is done by showing that $(u,t) \mapsto D^{W^R}_u B^\theta_t$ is jointly continuous.

Remark that the map $t \mapsto D^{W^R}_u B^\theta_t$ for $u \le t$ is given by (2.6.6) and hence it is continuous in time ($\forall t \in [u,T]$). Note now that Proposition 2.6.4 and Proposition A.3.1 yield that $D^{W^R}Z^{w,2}$ is bounded and $D^{W^R}_u Z^{w,2}_t = D^{W^R}_r Z^{w,2}_t = D^{W^R}_0 Z^{w,2}_t$ for any $0 \le u,r \le t \le T$. These properties hold as well for θ^R via the identity $-\gamma_R\theta^R = Z^{w,2}$.
Using the measure $\mathbb{P}^\theta$ (introduced in (2.3.3)), that $D^{W^R}\theta^S = 0$ and the identity $-\gamma_R\theta^R = Z^{w,2}$, one can rewrite (2.6.6) as

$$D^{W^R}_u B^\theta_t = D^{W^R}_u H^D + \frac{1}{\gamma_R}\int_t^T \kappa^R_s D^{W^R}_u Z^{w,2}_s \mathrm{d}s - \int_t^T \langle D^{W^R}_u \kappa^\theta_s, \mathrm{d}W^\theta_s \rangle. \quad (2.6.7)$$

[20] In other words, instead of a version we want κ^R_t to be a modification of $D^{W^R}_t B^\theta_t$.

Writing the same BSDE as above, but for a parameter v (instead of u) we have for $0 \leq v \leq t \leq T$

$$\begin{aligned} D_v^{W^R} B_t^\theta &= D_v^{W^R} H^D + \frac{1}{\gamma_R} \int_t^T \kappa_s^R D_v^{W^R} Z_s^{w,2} \mathrm{d}s - \int_t^T \langle D_v^{W^R} \kappa_s^\theta, \mathrm{d}W_s^\theta \rangle \\ &= D_u^{W^R} H^D + \frac{1}{\gamma_R} \int_t^T \kappa_s^R D_u^{W^R} Z_s^{w,2} \mathrm{d}s - \int_t^T \langle D_v^{W^R} \kappa_s^\theta, \mathrm{d}W_s^\theta \rangle, \end{aligned}$$

where we used the results of Proposition A.3.1 (part (iii)) and the above result that $D_u^{W^R} Z_t^{w,2} = D_r^{W^R} Z_t^{w,2} = D_0^{W^R} Z_t^{w,2}$ for any $0 \leq u, r \leq t \leq T$. Since the solution to (2.6.7) is unique and the BSDE just above has exactly the same parameters as (2.6.7), we must conclude that for any $t \in [0,T]$ and for $0 \leq u, r \leq t$ it holds $D_u^{W^R} B_t^\theta = D_r^{W^R} B_t^\theta$. From the continuity of $t \mapsto D_\cdot^{W^R} B_t^\theta$ follows now the joint continuity of $(u,t) \mapsto D_u^{W^R} B_t^\theta$ in its time parameters and hence the representation $D_t^{W^R} B_t^\theta = \kappa_t^R$ holds $\mathbb{P}$-a.s. for any $t \in [0,T]$. □

We can now prove Theorem 2.6.5.

Proof of Theorem 2.6.5. We proceed in the same way as in the proof of Proposition 2.6.4. The argument goes as follows: define the process $(e_t)_{t \in [0,T]}$ just like in (2.6.5); apply Itô's formula to $(e_t D_\cdot^{W^R} B_t^\theta)$ and write the resulting equation under $\mathbb{P}^\theta$ (just like (2.6.7)); take $\mathbb{P}^\theta$-conditional expectations. At this point a remaining Lebesgue integral is still in the dynamics:

$$\begin{aligned} D_u^{W^R} B_t^\theta &= (e_t)^{-1} \mathbb{E}^\theta \left[e_T D_u^{W^R} H^D + \frac{1}{\gamma_R} \int_t^T e_s (\kappa_s^R - D_u^{W^R} B_s^\theta) D_u^{W^R} Z_s^{w,2} \mathrm{d}s | \mathcal{F}_t \right] \\ &= (e_t)^{-1} \mathbb{E}^\theta \left[e_T D_u^{W^R} H^D | \mathcal{F}_t \right], \end{aligned}$$

where from the first to the second line we used Proposition 2.6.6, i.e. that $\kappa_s^R = D_s^{W^R} B_s^\theta = D_u^{W^R} B_s^\theta$ $\mathbb{P}$-a.s. for any $0 \leq u \leq s \leq T$.

Moreover, using Proposition A.3.1, and that the boundedness of $D^{W^R} Z^{w,2}$ implies that of e_T and e_t^{-1}, we conclude that the boundedness of $D_\cdot^{W^R} H^D$ implies that of $(D_\cdot^{W^R} B_t^\theta)$ and hence that of (κ_t^R).

Recalling $H^D = h^D(S_T, R_T)$ and the dynamics of R given by Equation (2.3.1), we see that (by the chain rule) $D_u^{W^R} H^D = b \partial_{x_2} h^D$. Since $b \partial_{x_2} h^D$ is either always positive or always negative and since $\kappa_t^R = D_t^{W^R} B_t^\theta$ $\mathbb{P}$-a.s. for any $t \in [0,T]$, it follows that $\kappa^R \neq 0$ $\mathbb{P}$-a.s. for all $t \in [0,T]$. More precisely, depending on the sign of $b \partial_{x_2} h^D$, κ^R is $\mathbb{P}$-a.s. either always positive or always negative[21], giving $\mathrm{sgn}(\kappa_\cdot^R) = \mathrm{sgn}(b \partial_{x_2} h^D)$. □

Remark 2.6.7 (Comparison with [HPDR10]). *There are some differences between this Section, in particular Subsections 2.6.1 and 2.6.2, and* [HPDR10, Section 4] *which we would like to point out. Starting from the setting and assumptions, as we consider entropic risk, whereas* [HPDR10] *switch to the problem of maximization of "entropic utilities" (i.e.,*

[21]For any positive random variable X ($X > 0$ $\mathbb{P}$-a.s.) one has $\mathbb{E}^{\mathbb{P}}[X|\mathcal{F}] > 0$ for any sigma-field $\mathcal{F}$. Since the measure change is done for a strictly positive density function, the inequality for the new conditional expectation is still strict.

expected exponential utility), there are some different signs within the BSDEs. Furthermore, θ_t^S is allowed to depend on the risk R_t, whereas we work in a Black-Scholes market, implying that θ^S is a constant.

Theorem 2.6.3, Proposition 2.6.4 (on Lipschitz property) and Theorem 2.6.5 (on market completeness) have a corresponding result in [HPDR10]. *However, our proofs of the latter two use purely probabilistic arguments relying on Malliavin calculus applied to the BSDEs for Y^w and B (in Proposition 2.6.6), whereas both Lipschitz continuity and market completeness are proven via PDE theory in* [HPDR10]. *The former approach appears to us a bit more straightforward than the latter, and it also yields slightly different results. For instance, with our approach we state the Markov property of Y^w and Z^w, not only that of the latter one, but we do not give the characterization as viscosity solution of a PDE.* [HPDR10] *show that $Z^{w,2}$ is a version of the Malliavin derivative of Y^w with respect to W^R, whereas our two-dimensional process Z^w is shown to be a modification of $D^{W^R}Y$. Furthermore, we give explicitly the regularity of B and κ and the BSDE solved by this couple of processes (in Proposition 2.6.6) .*

2.6.3 Parameter Analysis

It is possible to justify at a theoretical level some of the predictable behavior of the processes Y^w, B^θ and θ^R with relation to the problem's parameters n, γ_R, λ^a and γ_a for $a \in \mathbb{A}$. Some statements are only made for $t < T$, because the process at hand does not depend on a given parameter at terminal time T.

Theorem 2.6.8. *Let θ be the EMPR. The process (Y^w, Z^w) solving BSDE* (2.5.2) *is differentiable with relation to λ^a for any $a \in \mathbb{A}$, n and γ_R (see* (2.5.6) *and* (2.5.8)*).*
Fix agent $a \in \mathbb{A}$. If the differences

$$\gamma_R - \gamma_a \qquad \textit{and} \qquad \mathbb{E}^\theta\left[\Big(\sum_{b\in\mathbb{A}} w^b H^b\Big) - H^a\right] \tag{2.6.8}$$

are positive (negative respectively), then $\partial_{\tilde{\lambda}^a} Y_t^w$ is negative (positive respectively) for any $t \in [0,T]$. For any $a \in \mathbb{A}$ we have $\mathbb{P}$-a.s. that

$$\partial_{\gamma_R} Y_t^w < 0, \quad \partial_{\gamma_a} Y_t^w < 0, \qquad \forall t \in [0,T).$$

Furthermore, $\mathbb{P}$-a.s.

$$\partial_n Y_t^w < 0, \quad \textit{sgn}(\partial_n \theta_t^R) = \textit{sgn}(b\partial_{x_2} h^D), \forall t \in [0,T], \quad \textit{and} \quad \partial_n B_t^\theta < 0, \quad \forall t \in [0,T).$$

Part of the results are in some way expected. Introducing more derivatives leads to an overall risk reduction and as more derivatives are placed in the market, the derivative is worth less (per unit). If γ_R is interpreted as the representative agent's risk tolerance, then as γ_R increases we have a decrease in risk (Y^w decreases) since it represents an increase in the single agents' risk tolerance (i.e., γ_a increases).

The main message of the above theorem is that the effect of the performance concern of one agent on the aggregate risk depends essentially on how the agent is positioned with respect to the others, both in terms of risk tolerance as well as the personal endowments. If the agent's risk tolerance γ_a is higher than the aggregate risk tolerance γ_R and her

endowment position dominates the aggregate endowment position, then an increase in the agent's concern rate leads to an increase of the aggregate risk.
Before proving the above result we remark that condition (2.6.8) simplifies under certain conditions; such simplifications are summarized in the below corollary. All results follow by direct manipulation of the involved quantities.

Corollary 2.6.9. *Let the conditions of Theorem 2.6.8 hold.*
If $\gamma_a = \gamma$ for all $a \in \mathbb{A}$, then $\gamma_R - \gamma_a = \gamma\big(\sum_{a\in\mathbb{A}} w^a - 1\big) = 0$.
If $N = 2$ s.t. $\mathbb{A} = \{a, b\}$, then $w^b = 1 - w^a$ and hence

$$\Big(\sum_{c\in\mathbb{A}} w^c H^c\Big) - H^a = -w^b(H^a - H^b) \quad \textit{and} \quad \Big(\sum_{c\in\mathbb{A}} w^c H^c\Big) - H^b = w^a(H^a - H^b).$$

Similarly, $\gamma_R - \gamma_a = -w^b(\gamma_a - \gamma_b)$ and $\gamma_R - \gamma_b = w^a(\gamma_a - \gamma_b)$. Moreover, it holds that

$$\mathit{sgn}\Big(\partial_{\lambda^a} Y^w_t\Big) = -\mathit{sgn}\Big(\partial_{\lambda^b} Y^w_t\Big) \qquad \mathbb{P}\text{-a.s. for any } t \in [0,T]. \tag{2.6.9}$$

Proof of Theorem 2.6.8. Let $a \in \mathbb{A}$ and $t \in [0,T]$. Theorem 3.1.9 in [dR11] (see also [IDR10, Theorem 2.8]) ensures the differentiability of BSDE (2.5.2) with respect to γ_R, γ_a, λ^a and n.

$\triangleright$ *The derivative of Y^w in γ_R:* Applying ∂_{γ_R} to BSDE (2.5.2) and writing it under the probability measure $\mathbb{Q}$ defined in (2.6.3) yields the dynamics

$$\partial_{\gamma_R} Y^w_t = 0 + \int_t^T \left(-\frac{1}{2}(\theta^S_s)^2 - \frac{1}{2\gamma_R^2}\left(Z^{w,2}_s\right)^2 \right) \mathrm{d}s - \int_t^T \langle \partial_{\gamma_R} Z^w_s, \mathrm{d}W^{\mathbb{Q}}_s \rangle.$$

Taking $\mathbb{Q}$-conditional expectations and noticing that the Lebesgue integral term is strictly negative for any $t \in [0,T)$, we have then $\partial_{\gamma_R} Y^w_t < 0$ for any $t \in [0,T)$.

$\triangleright$ *The derivative of Y^w in γ_a:* This case follows from the previous one as γ_R is defined by (2.5.6) and the weights w^a (see (2.5.8)) are independent of γ_a for $a \in \mathbb{A}$. By definition $\gamma_R = \sum_{a\in\mathbb{A}} w^a \gamma_a$, which implies $\partial_{\gamma_a}\gamma_R = w^a > 0$. Finally, from $\partial_{\gamma_a} Y^w = \partial_{\gamma_R} Y^w \cdot \partial_{\gamma_a}\gamma_R$ the statement follows.

$\triangleright$ *The derivatives of Y^w in $\widetilde{\lambda}^a$:* We compute only the derivatives with respect to $\widetilde{\lambda}^a$ in order to present simplified calculations as $\widetilde{\lambda}^a = \lambda^a/(N-1)$. Calculating the involved derivatives leads to

$$\partial_{\widetilde{\lambda}^a} \frac{1}{1+\widetilde{\lambda}^a} = -(\Lambda w^a)^2, \quad \partial_{\widetilde{\lambda}^a} \frac{1}{\Lambda} = (w^a)^2, \quad \partial_{\widetilde{\lambda}^a} w^a = (w^a)^2\Lambda(w^a - 1), \quad \partial_{\widetilde{\lambda}^a} w^b = (w^a)^2\Lambda w^b,$$
$$\partial_{\widetilde{\lambda}^a}\gamma_R = \partial_{\widetilde{\lambda}^a} \sum_{b\in\mathbb{A}} w^b\gamma_b = (w^a)^2\Lambda(\gamma_R - \gamma_a) \quad \text{and} \quad \partial_{\widetilde{\lambda}^a} H^w = (w^a)^2\Lambda\Big(\Big(\sum_{b\in\mathbb{A}} w^b H^b\Big) - H^a\Big).$$

Combining the above results with the BSDE for $\partial_{\widetilde{\lambda}^a} Y^w$ under the $\mathbb{Q}$-measure (just as in the previous two steps) yields

$$\partial_{\widetilde{\lambda}^a} Y^w_t = -(w^a)^2\Lambda\mathbb{E}^{\mathbb{Q}}\left[\Big(\Big(\sum_{b\in\mathbb{A}} w^b H^b\Big) - H^a\Big) + (\gamma_R - \gamma_a)\int_t^T \left(\frac{1}{2}(\theta^S_s)^2 + \frac{1}{2\gamma_R^2}\left(Z^{w,2}_s\right)^2\right)\mathrm{d}s \Big| \mathcal{F}_t\right].$$

Since $\mathbb{Q}$ is equivalent to $\mathbb{P}$, the statement follows.

▷ *The derivative of Y^w in n:* Applying ∂_n to BSDE (2.5.2) and writing it under the probability measure $\mathbb{Q}$ defined in (2.6.3) yields the dynamics

$$\begin{aligned}
&\partial_n Y_t^w = \partial_n Y_T^w - \int_t^T \langle \partial_n Z_s^w, \mathrm{d}W_s^{\mathbb{Q}} \rangle \quad \text{with} \quad \partial_n Y_T^w = -\frac{H^D}{N} \\
\implies\ &\partial_n Y_t^w = \mathbb{E}^{\mathbb{Q}}[\partial_n Y_T^w | \mathcal{F}_t] = -\frac{1}{N}\,\mathbb{E}^{\mathbb{Q}}[H^D|\mathcal{F}_t] < 0,
\end{aligned}$$

where the last sign follows from the definition of Y_T^w and H^D.[22]

▷ *The derivative of θ^R in n:* The analysis of $Z^{w,2}$ and hence of θ^R with respect to n and γ_R follows from the analysis of (2.6.4). Given representation (2.5.11), applying ∂_n to BSDE (2.6.4) and writing it under the probability measure $\mathbb{Q}$ defined in (2.6.3) yields the dynamics

$$\begin{aligned}
&\partial_n D_u^{W^R} Y_t^w = \partial_n D_u^{W^R} Y_T^w - \int_t^T \langle \partial_n D_u^{W^R} Z_s^w, \mathrm{d}W_s^{\mathbb{Q}} \rangle + \int_t^T \left(\frac{1}{\gamma_R} D_u^{W^R} Z_s^{w,2} \partial_n Z_s^{w,2} \right) \mathrm{d}s \\
\iff\ &\partial_n Z_t^{w,2} = \partial_n D_u^{W^R} Y_T^w - \int_t^T \langle \partial_n D_u^{W^R} Z_s^w, \mathrm{d}W_s^{\mathbb{Q}} \rangle + \int_t^T \left(\frac{1}{\gamma_R} D_u^{W^R} Z_s^{w,2} \partial_n Z_s^{w,2} \right) \mathrm{d}s \\
\iff\ &\partial_n Z_t^{w,2} = (e_t)^{-1} \mathbb{E}^{\mathbb{Q}} \left[e_T \partial_n D_u^{W^R} Y_T^w | \mathcal{F}_t \right]
\end{aligned}$$

for $0 \le u \le t \le T$, where (e_t) is as in (2.6.5) and the argumentation is similar to that back there. Notice now that with $Y_T = -\sum_{a\in\mathbb{A}} w^a (H^a + nH^D/N)$ we have

$$\partial_n D_t^{W^R} Y_T^w = -\frac{1}{N} \left(\sum_{a\in\mathbb{A}} w^a \right) D_t^{W^R} H^D = -\frac{1}{N} \left(\sum_{a\in\mathbb{A}} w^a \right) b(\partial_{x_2} h^D)(S_T, R_T).$$

Although the index t on the LHS suggests otherwise, this expression is independent of t, which can be justified without appealing to the RHS by recalling (2.6.2).
Given Assumption 2.6.1, we conclude that $\mathrm{sgn}(Z_t^{w,2}) = -\mathrm{sgn}(b\partial_{x_2} h^D)$, and furthermore, recalling (2.5.11) and that $\gamma_R > 0$, we find that $\mathrm{sgn}(\partial_n \theta_t^R) = \mathrm{sgn}(b\partial_{x_2} h^D)$.

▷ *The derivative of B^θ in n:* We use justifications similar to those used in Proposition 2.6.6 and hence we do not give all the details. Recall (2.3.6), apply the ∂_n-operator to the equation and do the usual change of measure (with $\mathbb{P}^\theta$) to obtain

$$\partial_n B_t^\theta = 0 - \int_t^T \kappa_s^R \partial_n \theta_s^R \mathrm{d}s - \int_t^T \langle \partial_n \kappa_s^\theta, \mathrm{d}W_s^\theta \rangle \implies \partial_n B_t^\theta = -\mathbb{E}^\theta \left[\int_t^T \kappa_s^R \partial_n \theta_s^R \mathrm{d}s \right].$$

By the previous result we have $\mathrm{sgn}(\partial_n \theta_t^R) = \mathrm{sgn}(b\partial_{x_2} h^D)$ and from Theorem 2.6.5 we have $\mathrm{sgn}(\kappa_t^R) = \mathrm{sgn}(b\partial_{x_2} h^D)$. It easily follows that $\partial_n B_t^\theta < 0$. □

Unfortunately, the conditions used above do not allow for similar results on the behavior of, say $(\gamma_R, n, \lambda^\cdot) \mapsto \widetilde{Y}^a$. The conditions required for such results are too restrictive to be of any usefulness. Nonetheless, we will investigate them in Section 2.7 via numerical simulation.

[22] Observe that $\mathbb{Q}$ from (2.6.3) depends on n, hence $\mathbb{E}^{\mathbb{Q}}[H^D|\mathcal{F}_t]$ depends on n. Its sign, however, stays the same, as the conditional expectation of a positive random variable is always positive.

2.7 Study of a particular model with two agents

In this section, we investigate a model economy consisting of two agents using entropic risk measures and having opposed exposures to the external non-financial risk. We give particular attention to the impact of the relative performance concern rates on the equilibrium related processes. The model is simple enough to allow extended tractability, when compared with Sections 2.4, 2.5 and 2.6, and nonetheless still sufficiently general as to produce a rich set of results and interpretations. In particular, we explicitly describe the structure of the equilibrium. Using numerical simulations, we are able to explore the dependence of individual quantities (such as the optimal portfolios π^{*a} and minimized risks Y_0^a) on the various parameters, thus complementing the results in Theorem 2.6.8.

2.7.1 The particular model and numerical methodology

We consider a stylized market consisting of two agents. We argue that a larger set of N agents with certain exposures to the external risk R can be clustered in two groups: those profiting from the high values of R and those profiting from the low values of R, and we can apply the weighted aggregation technique used in Section 2.5 to each group. Our two agents can therefore be thought of as representative agents for each group. The external risk process is taken to be the temperature affecting the two agents, who also have access to a stock market.

Temperature and Stock models

We study one period of one month ($T = 1$). The temperature (in degrees Celsius) follows a stochastic differential equation (SDE) (2.3.1) with constant coefficients:

$$R_t = r_0 + \mu^R t + b\, W_t^R, \quad t \in [0, T],$$

and for the stock we take a standard Black–Scholes model:

$$\frac{dS_t}{S_t} = \mu^S dt + \sigma^S dW_t^S, \quad t \in [0, T],$$

where the coefficients are $R_0 = r_0 = 18$, $\mu^R = 2$ and $b = 4$ for the temperature process, and $S_0 = 50$, $\mu^S = -0.2$ and $\sigma^S = 0.25$ (so $\theta^S = \mu^S/\sigma^S = -0.8$) for the stock price process.

The agents' parameters and endowments and the derivative

Define $I(x) := \frac{1}{\pi}\arctan(x) + \frac{1}{2} \ \in [0, 1]$. The agents' endowments, H^a and H^b, are taken to be

$$\begin{aligned} H^a &= 5 + I\Big(2\big(R_T - 24\big)\Big) \cdot 15, \\ H^b &= 5 + I\Big(2\big(16 - R_T\big)\Big) \cdot \Big(15 + 5\, I\big(S_T - 40\big)\Big). \end{aligned}$$

Agent a profits from higher temperatures while agent b profits from lower ones. The derivative has a payoff H^D that does not depend on the stock S, and is given by

$$H^D = I(R_T - 20),$$

so that it allows to transfer the weather risk by trading the derivative. All functions satisfy Assumptions 2.3.1 and 2.6.1. Given the agents' opposite exposures to R_T and the design of H^D, agent a will act as a seller while agent b will act as a buyer, thus establishing a viable market for the derivative.

We assume throughout that the total supply of derivative is zero, $n = 0$; in other words, every unit of derivative one agent owns is underwritten by the other. The risk tolerance coefficients of the agents are fixed at $\gamma_a = \gamma_b = 1$ unless we are analyzing some behavior with respect to them. Similarly, unless otherwise specified, the concern rates are fixed to be $\lambda^a = \lambda^b = 0.25$.

The numerical procedure

The simulation of the involved processes entails a time discretization and Monte Carlo simulations. The forward processes have explicit solutions and we use them directly. All BSDEs are solved numerically. Regarding their time discretization, we use a standard backward Euler scheme, see [BT04], and we complement the time-discretization procedure with the control variate technique stated in [LdRS15, Section 5.4.2]. The approximation of the conditional expectations in the backward induction steps is done via projection over basis functions, see the Least-Squares Monte Carlo method used in [GT16].

We follow Sections 2.4 and 2.5. First, we solve the representative agent's BSDE (2.5.2). This yields via (2.5.11) the EMPeR process θ^R. Once this is obtained, we solve the BSDE for the price B^θ of the derivative, Equation (2.3.6), obtaining (κ^S, κ^R) in the process. Finally, we solve the BSDE (2.4.25) with driver (2.4.16) for each agent $a \in \mathbb{A}$ and compute the optimal strategies $\pi^{*,a} = (\pi^{*,a,1}, \pi^{*,a,2})$ via (2.4.17) and (2.4.18). We note that in the case of two agents, the system (2.4.17) is easily inverted.

All plots below are computed using 200.000 simulated paths along a uniform time-discretization grid of 20 time-steps, except the plot of Figure 2.2 which uses 30 time-steps.

2.7.2 Analysis of the behavior in the model

Figure 2.2 shows a realization of the behavior of the agents over the trading period. One can see that the price of the derivative reacts to the temperature's movements, in particular towards the end of the time period. This suggests that the derivative does indeed complete the market by providing the agents full exposure to R, or equivalently to W^R – Assumption 2.4.1 is satisfied. Agent b is always long in the derivative and a always short (the latter following from the former since her position is the opposite of that of b). The fact that both agents only go short in the stock is due to its decreasing trend ($\theta^S < 0$). Only agent b's terminal risk exposure depends directly on the terminal stock price, whereas only agent a has performance concerns[23].

[23] For this plot, $\lambda^b = 0$, hence agent b does not react to the trading performance of agent a.

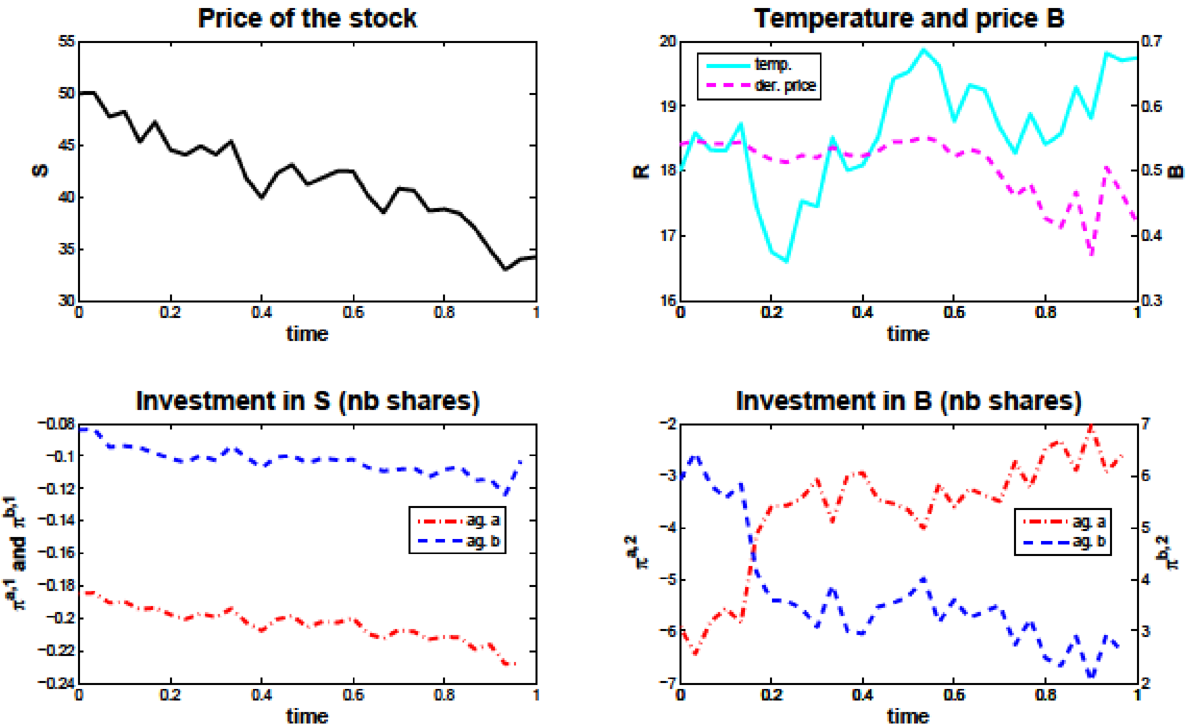


Figure 2.2: Sample paths of the relevant processes. Stock price on the top left; the temperature and the derivative price on the top right; the investment strategy in the stock on the bottom left and that in the derivative on the bottom right, for each agent. Here $\lambda^a = 0.25$ and $\lambda^b = 0.0$.

Trading activity

The optimal investment strategies for the derivative were seen in Remark 2.4.10 and are given by

$$\pi^{*,a,2} = \frac{1}{1+\lambda^a}\frac{\widetilde{Z}^{a,2} + \gamma_a \theta^R}{\kappa^R} \quad \text{and} \quad \pi^{*,b,2} = \frac{1}{1+\lambda^b}\frac{\widetilde{Z}^{b,2} + \gamma_b \theta^R}{\kappa^R}.$$

The optimal investment strategies in the stock follow easily by inverting A_2 from Expression (2.4.13). This yields

$$\begin{bmatrix} \pi^{*,a,1} \\ \\ \pi^{*,b,1} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{1-\lambda^a\lambda^b} & \dfrac{\lambda^a}{1-\lambda^a\lambda^b} \\ \dfrac{\lambda^b}{1-\lambda^a\lambda^b} & \dfrac{1}{1-\lambda^a\lambda^b} \end{bmatrix} \begin{bmatrix} \dfrac{\widetilde{Z}^{a,1} + \gamma_a \theta^S}{\sigma^S S} - \dfrac{\widetilde{Z}^{a,2} + \gamma_a \theta^R}{\kappa^R}\dfrac{\kappa^S}{\sigma^S S} \\ \dfrac{\widetilde{Z}^{b,1} + \gamma_b \theta^S}{\sigma^S S} - \dfrac{\widetilde{Z}^{b,2} + \gamma_b \theta^R}{\kappa^R}\dfrac{\kappa^S}{\sigma^S S} \end{bmatrix}. \qquad (2.7.1)$$

Remark 2.7.1 (On the structure of the equilibrium). *Let us comment on the structure of the investment strategy in stock.*

Each agent computes her strategy as a weighted sum of the way both would compute theirs if there was no relative performance concern (compare with (2.4.20)), using the weights $\left(\frac{1}{1-\lambda^a\lambda^b}, \frac{\lambda^a}{1-\lambda^a\lambda^b}\right)$ *for* a *and* $\left(\frac{\lambda^b}{1-\lambda^a\lambda^b}, \frac{1}{1-\lambda^a\lambda^b}\right)$ *for* b.

These weights can be understood from Equation (2.4.17) *with* $\mathbb{A} = \{a, b\}$*: each agent's best response is to invest in the stock according to her natural strategy plus* λ^i *times the strategy played by the other. Assume now that each agent was initially planning to compute her optimal position using*

$$\pi^{(0),i,1} = \frac{\widetilde{Z}^{i,1} + \gamma_i \theta^S}{\sigma^S S} - \frac{\widetilde{Z}^{i,2} + \gamma_i \theta^R}{\kappa^R} \frac{\kappa^S}{\sigma^S S}, \qquad i \in \{a, b\},$$

and that they are shown, in turn, the strategy that the other is about to play, so that they can update theirs, yielding a sequence of strategies $\pi^{(1),a,1}$*,* $\pi^{(1),b,1}$*,* $\pi^{(2),a,1}$*,* $\pi^{(2),b,1}$*,* $\pi^{(3),a,1}$*, … for each agent (starting with* a*'s update). Because they both update their strategy according to Equation* (2.4.17)*, we observe agent* a *imitating part of agent* b*, imitating part of agent* a*, imitating part of agent* b*, etc. Summing the corresponding series, agent* a *ends up investing according to* $\sum_{n=0}^{\infty} (\lambda^a \lambda^b)^n \, \pi^{(0),a,1} + \lambda^a \sum_{n=0}^{\infty} (\lambda^a \lambda^b)^n \pi^{(0),b,1}$*, and similarly for agent* b*. This geometric series converges to* (2.7.1)*.*

The structure of the optimal investment in the derivative, however, follows from the endogenous trading condition. If an agent is shown the strategy that the other had decided to follow, she could not unilaterally change her strategy. From this emerges the EMPeR θ^R *– see below.*

We now look at the behavior of the individual portfolios with respect to the rates of relative performance concern. The intensity of the trading activity at time $t = 0$ on both the stock ($\pi_0^{*,a,1}$) and the derivative ($\pi_0^{*,a,2}$) as maps of the concern rates λ^a, λ^b can be found in Figure 2.3. The positions of agent b are similar in some sense: for the stock, the surface looks very similar; for the derivative, it is the exact opposite (due to the zero net supply condition). For readability we plot only the position of agent a.

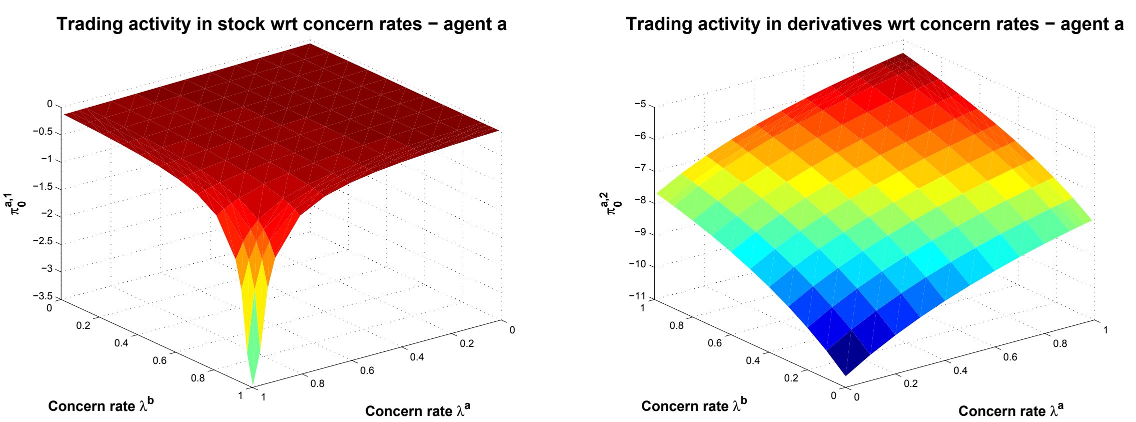


Figure 2.3: Initial number $\pi_0^{a,1}$ and $\pi_0^{a,2}$ of shares of stock (left) and derivative (right) held by agent a, as a function of (λ^a, λ^b). For visualization purposes the axes on the left picture were inverted.

The observed behavior in Figure 2.3 is in line with the intuitive idea that the more the agents are concerned (high λ^i) about their relative performance $V_T^i - V_T^j$, $j \neq i \in \{a, b\}$ (recall (0.0.1)), the more they will invest in a way that neutralizes this source of risk. This is done by adopting a trading strategy that is as close as possible to that of the other agent.

For the stock, we see from the formulas in Remark 2.7.1 that when $\lambda^a\lambda^b < 1$, the process of a imitating b imitating a, etc, results in a finite position. But the volume increases with both λ^a and λ^b, and explodes as $(\lambda^a, \lambda^b) \to (1, 1)$. In our example they would both (short-)sell infinitely many shares of the stock. Note that this is possible only because the stock is assumed to be exogenously priced and perfectly liquid. For the derivative, they cannot imitate each other and position themselves in the same direction, as the zero net supply condition implies that the agents must hold exactly opposite positions. Agent b's gains on trading the derivative will be exactly agent a's losses. The only way to reduce the difference in performances for a very concerned agent is to engage less (in volume) in the trading of the derivative. The market clearing condition then forces the other agent to also trade less (in volume). This is seen from the factor $1/(1 + \lambda^i)$ in the formulas in Remark 2.7.1 and is confirmed in Figure 2.3 (on the right) where agent a, identified as the seller, ends up selling fewer units of the derivative as either concern rate increases. Due to the market clearing condition between the agents, no explosion is possible.

Price of the derivative

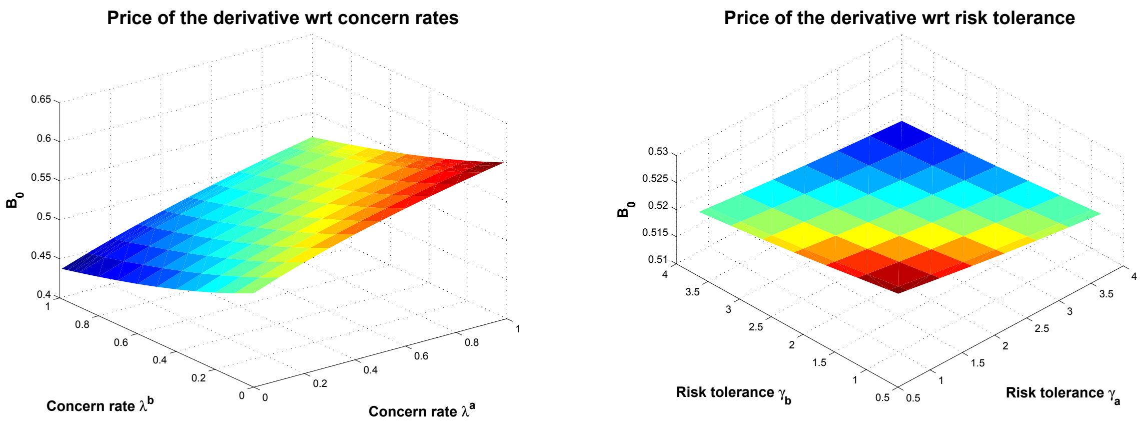


Figure 2.4: Initial price of the derivative, B_0^θ, as map of (λ^a, λ^b) on the left, and as map of (γ_a, γ_b) on the right.

Figure 2.4 (left) shows an opposite dependence of the derivative's price B_0^θ on the concern rates λ^a, λ^b, a behavior not captured by Theorem 2.6.8. One can make sense of this effect by having in mind Figure 2.3. A higher λ^a implies that agent a wants to trade less and, as she is the seller, this drives the price up. Symmetrically, a higher λ^b implies that agent b wants to trade less and, as she is the buyer, this drives the price down. Figure 2.4 (right) shows a negative relation between risk tolerance and the price of the derivative. As higher risk tolerance requires less risk hedging, the demand for the derivative decreases, which leads to the observed lower prices for the derivative.

Aggregated risk

Figure 2.5 confirms the analytical results of Theorem 2.6.8 and Corollary 2.6.9. As predicted, the increase of the risk tolerances leads to a decrease in the aggregated risk (see Figure 2.5, right picture). The picture on the left shows clearly the cross behavior stated in (2.6.9).

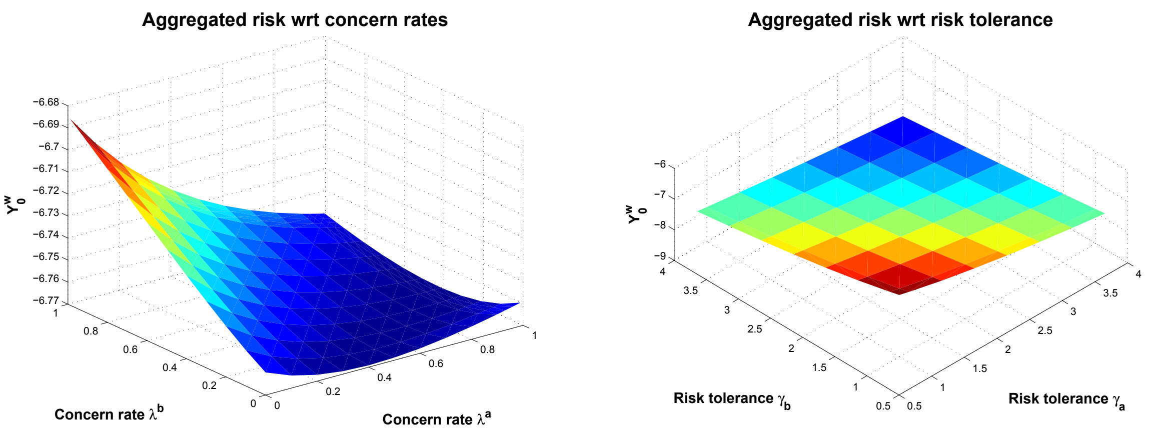


Figure 2.5: Aggregated risk Y_0^w as a function of (λ^a, λ^b) (left) and of (γ_a, γ_b) (right).

Risk of each agent

Theorem 2.6.8 does not capture the behavior of each agent's risk assessment as a function of the concern rates $\lambda^{\cdot}$.

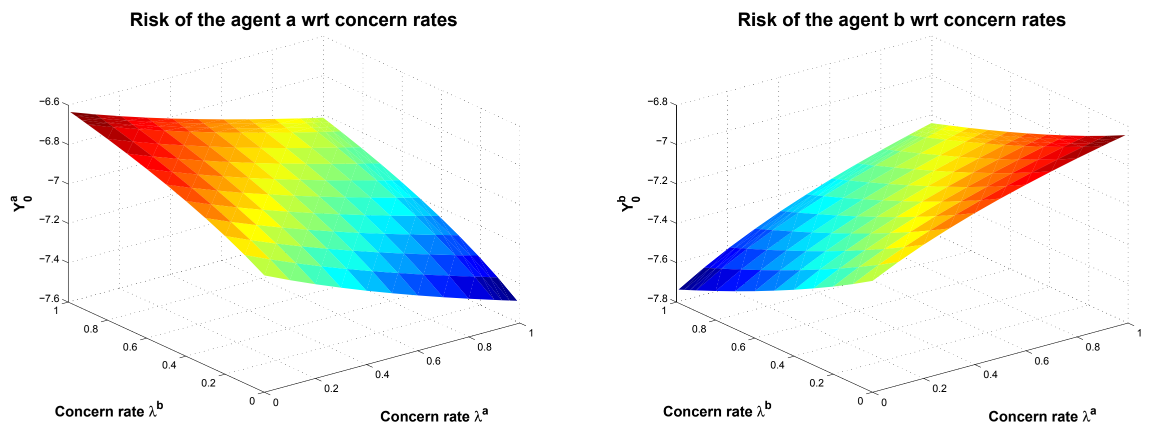


Figure 2.6: Risk Y_0^a (left) and Y_0^b (right) as a function of (λ^a, λ^b).

Figure 2.6 portrays the risk perceptions of each agent as λ^a, λ^b change. First, agent a's risk Y_0^a increases in λ^b. This can be explained as follows: As λ^b increases, agent b engages in less trading of the derivative in order to reduce her relative performance concern, and this affects agents a's ability to hedge H^a. Second, agent a's risk Y_0^a decreases in λ^a. A possible explanation for this behavior is, having in mind (0.0.1) or (2.3.10), that if a gives more importance to her relative performance concern, then she trades more in a way that mimics what b does. This in turn reduces her ability to neutralize the endowment risk.

2.7.3 Effect of introducing the derivative

We now comment on the effects of introducing the derivative in this model market. Figure 2.7 displays the risks of the representative agent and of agent a with respect to λ^a and λ^b when no derivative is available and when a market for it is available.

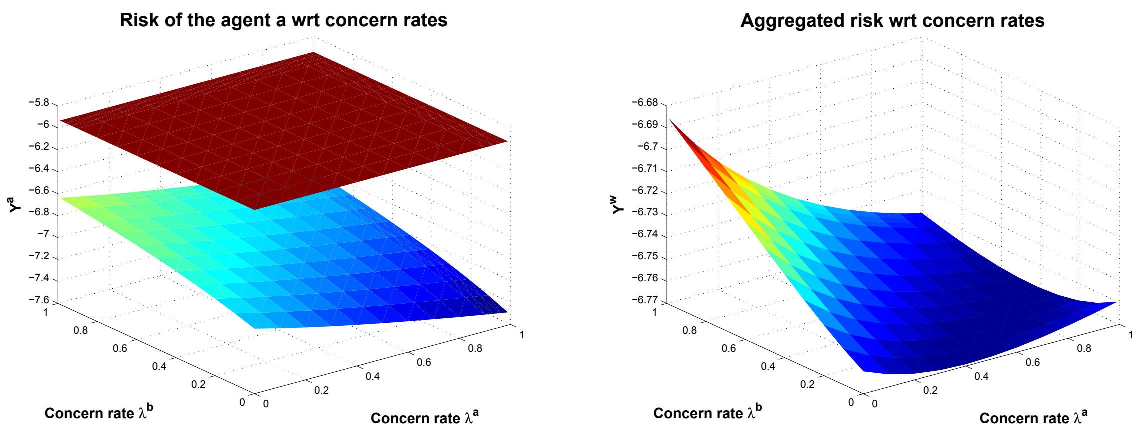


Figure 2.7: Left : risk Y_0^a when the derivative is not available (flat surface) and when it is (tilted surface), as a function of the concern rates (λ^a, λ^b). Right: same plot for the aggregated risk Y_0^w (the two surfaces are equal).

We observe in the plot on the right that adding the derivative does not change the aggregated risk. This is clear if one views it as the risk of the representative agent: being alone by construction, the zero net supply condition means that she must keep a zero position in the derivative, and hence does not benefit from its presence (compare the strategies in (2.4.22) with those in Example 2.5.6).
For an individual agent however (left plot), the availability of the derivative always leads to a reduction of risk. We observe that in the absence of the derivative, the risk of agent a does not depend on the concern rates. We can apply the methodology of Sections 2.4 and 2.5 to find that the optimal porfolios of the agents in this situation are given by

$$\pi_t^{*,i,1} = \frac{1}{1-\lambda^a\lambda^b}\,\frac{\gamma_i\theta^S + \widetilde{Z}_t^{i,1}}{\sigma^S S_t} + \frac{\lambda^i}{1-\lambda^a\lambda^b}\,\frac{\gamma_j\theta^S + \widetilde{Z}_t^{j,1}}{\sigma^S S_t} \quad \text{for } j \neq i \in \{a,b\},\ t \in [0,T],$$

while the minimized risk equation is given by the BSDE

$$\begin{aligned} d\widetilde{Y}_t^i &= -\Big[-\frac{1}{2}\gamma_i\big(\theta^S\big)^2 - \widetilde{Z}_t^{i,1}\theta^S + \frac{1}{2\gamma_i}\big(\widetilde{Z}^{i,2}\big)^2\Big]\mathrm{d}t + \langle \widetilde{Z}_t^i, \mathrm{d}W_t\rangle, \quad t \in [0,T] \\ \widetilde{Y}_T^i &= -H^i(S_T, R_T). \end{aligned}$$

This shows analytically that the value of the problem, Y_0^a, depends on neither λ^a nor λ^b while the optimal strategy does, as was already observed in [FdR11, Proposition 4.1].

Playing the game repeatedly leads to an extreme impact on the stock market

The above study considers a one-period model with (continuous-time) trading until the horizon $T = 1$ month. Imagine now the repetition of this trading period over time and assume no significant changes to the agents' endowments or the dynamics of the financial and external risks.

At the level of the agents' preferences, with the sole exception of the concern rates, we assume that they do not change with time. Specifically, we assume that their risk tolerances, and consequently the entropic risk measures $\rho_0^{\cdot}$ used to assess their risk in (0.0.1), are fixed throughout; however, their concern rates $\lambda^{\cdot}$ over their relative performance may vary. This might happen in an increasingly competitive segment where the participants feel an increasing pressure to surpass their competitors in order to survive. Figure 2.6 sheds some light on the outcome of playing this game repeatedly. Indeed, each agent benefits from a unilateral increase of their concern rate λ while they are worse off with an increase of the other's concern rate. So they have an incentive to increase λ, as the trading periods are repeated, culminating in Assumption 2.3.3 being violated as $(\lambda^a, \lambda^b) \to (1, 1)$.

It is interesting to note that this drifting toward the singularity of the model, $(\lambda^a, \lambda^b) = (1, 1)$, is not captured by the risk assessments. Figures 2.5 and 2.6 show that Y_0^w, Y_0^a and Y_0^b remain bounded. At the level of the investment strategies, the trading activity in the derivative slows down but persists. The sharing of the external risk becomes less efficient, because the agents are increasingly concerned about losing out to the other, but does not disappear. However, the investment in the stock explodes (see Figure 2.3). We stress that this behavior arises only *after* the derivative is introduced in the market. Indeed, as shown by Figure 2.7, when the derivative is not available and the agents in $\mathbb{A}$ are only concerned with the relative performance of their investment strategy on the stock market, they have no incentive to have increasingly high concern rates. The particular shape of the surface $(\lambda^a, \lambda^b) \mapsto Y_0^i$, risk decreasing with λ^i but increasing with λ^j, appears only when the derivative is made available. In this situation, the agents are placed in direct interaction (by trading) in addition to the indirect one (social): each agent makes now gains directly *over the other*. The final result is a potential destabilization of the stock market.

2.8 Conclusion

In this chapter, we analyzed the effect of a form of social interaction between agents on an equilibrium pricing mechanism. Specifically, we considered the pricing of a (market-completing) derivative introduced to allow market participants to share the risk associated with an external and non-tradable risk factor. The social interaction here takes the form of concerns over relative performance.

From a theoretical point of view, we have shown how to solve the problem for general risk measures and a finite number of agents, when assuming that the derivative completes the market. Due to the heterogeneous rates of concerns of the agents, the risks of the agents cannot be aggregated by the usual infimal convolution technique, so we developed it further and introduced the *weighted-dilated infimal convolution* variant.

We then focused on the particular case of the entropic risk measure and were able to determine sufficient conditions to design a derivative that completes the market. In a market model with two agents representing opposite profiles of exposure to the external risk, we explored the impact of the social interactions on the benefit brought by financial innovation.

We found that the introduction of the derivative always reduces the risk, at the level of individual agents. However, the particular distribution of this risk reduction means that both agents have an incentive to become more concerned with their relative performance, because this reduces their individual risk. At the global level, while the volume of derivatives traded merely decreases, the volumes traded in the previously-existing financial asset explode and the equilibrium ceases to exist. This non-trivial, and perhaps not desirable, behavior of the system after introduction of the derivative bears resemblance to the findings of [CMV09] and [CML16]. In practice, the assumption that the agents are small and that the price dynamics of the stock is independent of their actions fails to hold. Thus, although the stock price is fundamentally independent of the external risk, introducing the derivative can lead to unintended consequences on what was a stable stock market. We stress that this phenomenon is not captured by the risk measures. Therefore, one should not only use the performance of the risk measure when evaluating the possible benefits of a new policy (the introduction of the derivative, here). This also stresses the importance of having systemic view: studying the problem from the point of view of an individual investor shows that the availability of the derivative is always beneficial, but at the global level the picture has strong nuances. Strongly undesirable endogenous phenomena can emerge in the dynamics, arising essentially from the interaction between the various agents and their possibility to adapt to the new policy.

3 Pathwise Directional Derivatives Beyond Cameron-Martin Directions

3.1 Organization of this chapter

First, in Section 3.2, the spaces and necessary notation are introduced. In Section 3.3 variations with Cameron-Martin functions are presented as benchmark for our further analysis. In Sections 3.4 and 3.5 we extend this to variations from the space of Hölder continuous and merely continuous functions. Section 3.6 goes another step further to discontinuous functions and Section 3.7 puts forward a notion of variation of paths with measures. Finally, Section 3.8 concludes with a table of the relevant results.

3.2 Spaces and Notation

Let $I := [0,1]$ denote our time interval and $\Omega := C(I;\mathbb{R})$ the space of real-valued continuous paths equipped with the supremum norm $\|f\|_\infty := \sup\{|f(t)| \mid t \in I\}$, where $|\cdot|$ shall denote the Euclidean norm on $\mathbb{R}$. By $\mathfrak{B}(I)$ we denote the Borel σ-algebra on I, whereas $\mathfrak{B} = \mathfrak{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$. Let H denote some Hilbert space on I, for instance $H = L^2(I) = L^2(I, \mathfrak{B}(I), \lambda|_I)$, with a fixed ONB[1] $(h_k)_{k\in\mathbb{N}}$. Denote by $C_b^\infty(\mathbb{R}^n)$ the space of bounded and infinitely continuously differentiable functions on $\mathbb{R}^n$ with bounded partial derivatives. Let $\mathcal{S}$ denote what we will call cylinder functions, i.e.

$$\mathcal{S} := \{F\colon \Omega \to \mathbb{R} \mid F(\omega) = f(\theta_1(\omega), \ldots, \theta_n(\omega)), f \in C_b^\infty(\mathbb{R}^n), n \in \mathbb{N}\},$$

where $\theta_k(\omega) := \int_I h_k(t)\mathrm{d}\omega(t)$. If we want to emphasize the basis used we also write $\theta_k^h(\omega)$.

Remark 3.2.1. *If $(h_k)_{k\in\mathbb{N}}$ is an ONB of $L^2(I)$, then in general the integral $\theta_k^h(\omega) = \int_I h_k(t)\mathrm{d}\omega(t)$ need not exist for all $\omega \in \Omega$. Instead, F could have the representation in $\mathcal{S}$ only for almost all $\omega \in \Omega$. As we work pathwise throughout this section, all we require is that for the fixed ω, $\theta_k^h(\omega)$ exists. For the specific basis of Haar functions, which we introduce in Section 3.4.1 and use thereafter, the integrals do exist.*

We also work with functions with infinite representations, i.e. functions belonging to

$$\mathcal{D} := \left\{F\colon \Omega \to \mathbb{R} \,\middle|\, F(\omega) = f\left((\theta_k(\omega))_{k\in\mathbb{N}}\right),\ f \in C^\infty(\mathbb{R}^{\mathbb{N}}),\ (\beta_k(\omega))_{k\in\mathbb{N}} \in \ell^2\right\},$$

[1] We use the often used wording *orthonormal basis*, by which we mean a complete orthonormal system.

where for $k \in \mathbb{N}$

$$\theta_k(\omega) := \int_I h_k(t)\mathrm{d}\omega(t) \quad \text{and} \quad \beta_k(\omega) := \nabla_k f\left((\theta_n(\omega))_{n\in\mathbb{N}}\right).$$

The Cameron-Martin space on I is defined as

$$\mathcal{H}(I) := \left\{ f\colon I \to \mathbb{R} \;\middle|\; \exists g \in L^2(I)\colon f(t) = \int_0^t g(s)\mathrm{d}s,\, t \in I \right\}.$$

The α-Hölder space on I is defined as

$$C^\alpha(I;\mathbb{R}^d) := \left\{ f\colon I \subset \mathbb{R} \to \mathbb{R}^d \;\middle|\; \exists C > 0\colon \|f(t) - f(s)\| \le C|t-s|^\alpha,\ \forall t, s \in I \right\}$$

for some $0 < \alpha \le 1$, where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^d$. We restrict ourselves here to $d = 1$ for better readability, hence we write $C^\alpha(I) := C^\alpha(I;\mathbb{R})$. When we restrict the analysis to functions from a space vanishing at zero, we denote this by adding subscript 0, e.g. $C_0(I) = \{ f \in C(I) \mid f(0) = 0\}$. The Hölder coefficient[2] of f on I is

$$|f|_\alpha := \sup\left\{ \frac{|f(t) - f(s)|}{|t-s|^\alpha} \;\middle|\; s, t \in I,\, s \ne t \right\}$$

such that $C_0^\alpha(I) = \{ f\colon I \to \mathbb{R} \mid f(0) = 0,\, |f|_\alpha < \infty\}$. A norm on $C^\alpha(I)$ is given by $\|f\|_\alpha := \|f\|_\infty + |f|_\alpha$.

For $p \ge 1$ let $V_p(f)$ denote the p-variation of a function f on I and let $\mathcal{V}_p(I)$ denote the space of functions on I with finite p-variation. For the exact definition and properties of this space, see Appendix B.2.4.

When we work with measures they shall belong to the space of finite measures on $(I, \mathfrak{B}(I))$, which we denote by $\mathfrak{M}(I)$, and the space of measurable functions will be denoted by

$$\mathcal{M}(I) := \{ f\colon (I, \mathfrak{B}(I)) \to (\mathbb{R}, \mathfrak{B}(\mathbb{R})) \mid f\ \mathfrak{B}(I) - \mathfrak{B}(\mathbb{R}) - \text{measurable}\}.$$

The Dirac measure on $\mathfrak{B}(I)$ at $t \in I$,

$$\delta_t(A) := \mathbb{1}_A(t) = \begin{cases} 1 & , t \in A, \\ 0 & , t \notin A \end{cases} \qquad \text{for } A \in \mathfrak{B}(I),$$

shall not be confused with the Kronecker delta

$$\delta_{ab} := \begin{cases} 1 & , a = b, \\ 0 & , a \ne b \end{cases} \qquad \text{for } a, b \in \mathbb{R},$$

which always has two subscripts, but no argument. Both are of course linked via the identity $\delta_{ab} = \delta_a(\{b\}) = \delta_b(\{a\})$ for $a, b \in I$.

In $L^2(I) = \left\{ f\colon I \to \mathbb{R}^d \;\middle|\; \|f\|_{L^2(I)} < \infty \right\}$ with $\|f\|_{L^2(I)} := \left(\int_I |f(t)|^2\,\mathrm{d}t\right)^{1/2}$ we take the usual inner product $\langle f, g\rangle_{L^2(I)} := \int_I f(t)g(t)\mathrm{d}t$. If f is integrable w.r.t. g, we write $\langle f, \mathrm{d}g\rangle := \int_I f(t)\mathrm{d}g(t)$. Observe that if $g \in \mathcal{H}(I)$, then $\langle f, \mathrm{d}g\rangle = \langle f, \dot{g}\rangle_{L^2(I)}$.

[2]The Hölder coefficient is not a norm on $C^\alpha(I)$ as can be seen when applying it to non-zero constant functions. On the subspace $C_0^\alpha(I)$, however, it is a true norm.

We also recall the following sequence spaces:

$$\ell^p := \left\{ x = (x_n)_{n\in\mathbb{N}} \subset \mathbb{R} \,\middle|\, \|x\|_{\ell^p} = \left(\sum_{n\in\mathbb{N}} |x_n|^p \right)^{1/p} < \infty \right\} \text{ for } p \in [1,\infty),$$
$$\ell^\infty := \left\{ x = (x_n)_{n\in\mathbb{N}} \subset \mathbb{R} \,\middle|\, \|x\|_{\ell^\infty} = \sup_{n\in\mathbb{N}} |x_n| < \infty \right\}.$$

A collection of results from sequence spaces is presented in Appendix B.2.3.

We use the notation $\nabla_k f := \frac{\partial f}{\partial x_k}$ for the k-th partial derivative of f and let $\nabla f := (\nabla_k f)_k$ denote the (possibly infinite dimensional) gradient of f.

The spaces of Riemann-, Lebesgue- and generalized Riemann-integrable functions (on I) will be denoted by $\mathcal{R}(I)$, $\mathcal{L}(I)$ and $\mathcal{G}(I)$, respectively. By $f \in \mathcal{R}(g)$ (resp. $\mathcal{L}(g)$ or $\mathcal{G}(g)$) we denote that f is RS (resp. LS or generalized RS) integrable w.r.t. g on I. We write $(\mathcal{R}) \int_I f(t)\mathrm{d}g(t)$ (resp. $(\mathcal{L}) \int_I f(t)\mathrm{d}g(t)$ or $(\mathcal{G}) \int_I f(t)\mathrm{d}g(t)$) if we want to emphasize the type of the Stieltjes integral. Further details on the different types of integration can be found in Appendix B.1.

The space of regulated functions on I (taking values in $\mathbb{R}$) shall be denoted by $\mathfrak{R}(I)$. The definition of a regulated function and properties are given in Appendix B.2.2.

3.3 Varying paths with in Cameron-Martin functions

The starting point of our analysis is the directional derivative of cylinder functions $F \in \mathcal{S}$ that is obtained by varying paths $\omega \in \Omega$ by functions in $\mathcal{H}(I)$, the Cameron-Martin space.[3] This is extended in a straightforward manner to functions $F \in \mathcal{D}$, and different aspects of our setting and the directional derivatives are discussed. This section's results serve as a benchmark for the extensions in later sections by providing results and methodology. Until further notice, all integrals are of Lebesgue- or Lebesgue-Stieltjes type and they are always integrals over time, namely the fixed interval $I = [0,1]$. Throughout this section we also fix an ONB $(h_k)_{k\in\mathbb{N}}$ of $L^2(I)$ such that for a fixed $\omega \in \Omega$ the integrals $\theta_k^h(\omega)$ exist.

3.3.1 Cylinder functions

We fix a cylinder function $F \in \mathcal{S}$ that has the representation $F(\omega) = f(\theta_1(\omega), \dots, \theta_n(\omega))$ with $\theta_k(\omega) := \int_I h_k(t)\mathrm{d}\omega(t)$ $(k = 1, \dots, n)$. Further let $\gamma = \int_I \dot{\gamma}(t)\mathrm{d}t \in \mathcal{H}(I)$ with

[3] As a reference for this section on may consult for example [Nua06].

$\dot{\gamma} \in L^2(I)$. Then[4]

$$\begin{aligned}
D_\gamma F(\omega) :&= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left[F(\omega+\varepsilon\gamma) - F(\omega)\right] \\
&= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\Bigg[f\left(\int_I h_1 \mathrm{d}(\omega+\varepsilon\gamma), \dots, \int_I h_n \mathrm{d}(\omega+\varepsilon\gamma)\right) \\
&\qquad - f\left(\int_I h_1\mathrm{d}\omega, \dots, \int_I h_n \mathrm{d}\omega\right)\Bigg] \\
&= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\Bigg[f\bigg(\int_I h_1\mathrm{d}\omega + \varepsilon\int_I h_1 \mathrm{d}\gamma, \dots, \int_I h_n\mathrm{d}\omega + \varepsilon \int_I h_n\mathrm{d}\gamma\bigg) \\
&\qquad - f\left(\int_I h_1\mathrm{d}\omega, \dots, \int_I h_n \mathrm{d}\omega\right)\Bigg] \\
&= \sum_{k=1}^n \frac{\partial f}{\partial x_k}\left(\int_I h_1\mathrm{d}\omega, \dots, \int_I h_n\mathrm{d}\omega\right)\int_I h_k \mathrm{d}\gamma \\
&= \sum_{k=1}^n \nabla_k f\left(\theta_1(\omega), \dots, \theta_n(\omega)\right)\langle h_k, \dot{\gamma}\rangle_{L^2(I)}.
\end{aligned} \tag{3.3.1}$$

The integral $\int_I h_k(t)\mathrm{d}\gamma(t) = \int_I h_k(t)\dot{\gamma}(t)\mathrm{d}t$ exists for $h_k \in \mathcal{H}(I)$ and $\dot{\gamma} \in L^2(I)$, as this is the usual scalar product in $L^2(I)$. If $F \in \mathcal{S}$, then $f \in C^\infty$, hence all partial derivatives exist. The sum is finite, hence $D_\gamma F(\omega)$ is well-defined as pathwise limit (in $(\mathbb{R}, |\cdot|)$) for all $\gamma \in \mathcal{H}(I)$.

Remark 3.3.1. *An important property that we used here is that for $k \in \{1, \dots, n\}$, the mapping $\theta_k\colon \Omega \to \mathbb{R}$, given by $\theta_k(\omega) = \int_I h_k(t)\mathrm{d}\omega(t)$, is linear, i.e., for $\omega, \widehat{\omega} \in \Omega$ and $\varepsilon \in \mathbb{R}$ we have $\theta_k(\omega + \varepsilon\widehat{\omega}) = \theta_k(\omega) + \varepsilon\theta_k(\widehat{\omega})$. It is the bilinearity of the Stieltjes integral $(f, g) \mapsto \int_I f\mathrm{d}g$ that we use here. See for instance II.11.1 for the bilinearity of the RS integral and VII.11.11-12 for that of the LS integral (both in* [Hil63]*). Even the rough integral is bilinear, see for example Lemma 2.6 in* [HW13] *or Theorem 1 in* [Gub04].

3.3.2 From cylinder functions to infinite representation

Recall the space of functions we want to work with:

$$\mathcal{D} := \left\{ F\colon \Omega \to \mathbb{R} \,\middle|\, F(\omega) = f\left((\theta_k(\omega))_{k\in\mathbb{N}}\right),\ f \in C^\infty(\mathbb{R}^\mathbb{N}),\ \left(\beta_k(\omega)\right)_{k\in\mathbb{N}} \in \ell^2\right\},$$

where for $k \in \mathbb{N}$

$$\theta_k(\omega) := \int_I h_k(t)\mathrm{d}\omega(t) \quad \text{and} \quad \beta_k(\omega) := \nabla_k f\left((\theta_n(\omega))_{n\in\mathbb{N}}\right).$$

The assumption that the partial derivatives form a sequence in ℓ^2 is natural in the context of Malliavin calculus, see e.g. Theorem 1.5.2 or the corresponding observation above Definition 2.2.1 in [Imk08].

[4]Throughout we use the symbol $D_\gamma F(\omega)$ to denote $\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left[F(\omega+\varepsilon\gamma) - F(\omega)\right]$ if the latter should exist.

If the limit exists, then we can define for $F \in \mathcal{D}$ and $\gamma \in \mathcal{H}(I)$ the directional derivative of F in ω in direction of γ:

$$
\begin{aligned}
D_\gamma F(\omega) :&= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left[F(\omega + \varepsilon\gamma) - F(\omega) \right] \\
&= \sum_{k=1}^{\infty} \nabla_k f \left(\left(\int_I h_n \mathrm{d}\omega \right)_{n \in \mathbb{N}} \right) \int_I h_k \mathrm{d}\gamma \\
&= \sum_{k=1}^{\infty} \beta_k(\omega) \langle h_k, \dot{\gamma} \rangle_{L^2(I)}.
\end{aligned}
\tag{3.3.2}
$$

As $(h_k)_{k\in\mathbb{N}}$ form an ONB of $L^2(I)$, the sequence $(\langle h_k, \dot{\gamma}\rangle)_{k\in\mathbb{N}}$ lies in ℓ^2. This can be seen from Parseval's identity, which states that if $(h_k)_{k\in\mathbb{N}}$ is an ONB of $L^2(I)$, then for all $f \in L^2(I)$ one has

$$
\sum_{k\in\mathbb{N}} |\langle f, h_k \rangle|^2 = \|f\|^2_{L^2(I)} .
$$

Furthermore, by assumption $\beta_k(\omega) = \nabla_k f\left((\theta_n(\omega))_{n\in\mathbb{N}}\right)$ $(k \in \mathbb{N})$ defines a sequence in ℓ^2. This shows (using Lemma B.2.7), that the summands in (3.3.2) form a sequence in ℓ^1, i.e., the series converges (absolutely). We can summarize these thoughts as follows:

Theorem 3.3.2. *Let $F \in \mathcal{D}$ with the above notation, fix $\omega \in \Omega$.*
Then $D_\gamma F(\omega) = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left[F(\omega+\varepsilon\gamma) - F(\omega)\right]$ is given by the absolutely convergent series $\sum_{k=1}^{\infty} \beta_k(\omega)\langle h_k, \dot{\gamma}\rangle_{L^2(I)}$.

3.3.3 Dependence on the ONB chosen

We defined $D_\gamma F(\omega)$ for a given $\omega \in \Omega$ and a fixed ONB of $H = L^2(I)$. One may wonder whether the object we defined is basis-independent.[5]

Let $(g_k)_{k\in\mathbb{N}}$ and $(h_k)_{k\in\mathbb{N}}$ be two different orthonormal bases of $L^2(I)$ such that $\theta_k^g(\omega)$ and $\theta_k^h(\omega)$ exist. With these fixed, assume that $F(\omega)$ has the representations

$$
F(\omega) = f^h \left(\int_I h_1(t)\mathrm{d}\omega(t), \ldots, \int_I h_n(t)\mathrm{d}\omega(t) \right) = f^g \left(\int_I g_1(t)\mathrm{d}\omega(t), \ldots, \int_I g_m(t)\mathrm{d}\omega(t) \right)
$$

for two possibly different numbers $n, m \in \mathbb{N}$ and $f^h \in C_b^\infty(\mathbb{R}^n;\mathbb{R})$ and $f^g \in C_b^\infty(\mathbb{R}^m;\mathbb{R})$. Assume further that the linear hulls of $h_1, \ldots, h_n$ and $g_1, \ldots, g_m$ are identical. Then we have $h_i = \sum_{j=1}^m \langle h_i, g_j\rangle g_j$ for $i = 1, \ldots, n$. Define the matrix

$$
M := \left(\langle h_i, g_j \rangle\right)_{1\le i\le n, 1\le j\le m} .
$$

With this we see that $f^h \circ M = f^g$, hence with the notation $\theta_k^h(\omega) := \int_I h_k(t)\mathrm{d}\omega(t)$ and $\theta_k^g(\omega) := \int_I g_k(t)\mathrm{d}\omega(t)$ for $k \in \mathbb{N}$ we find that

[5]We give the arguments as they are presented on page 42 in [Imk08] for completeness of the exposition and add remarks concerning the finiteness of the representation.

$$\begin{aligned}\sum_{j=1}^{m} \nabla_j f^g(\theta_1^g(\omega),\dots,\theta_m^g(\omega))g_j &= \sum_{j=1}^{m} \nabla_j (f^h \circ M)(\theta_1^g(\omega),\dots,\theta_m^g(\omega))g_j \\ &= \sum_{j=1}^{m}\sum_{i=1}^{n} \nabla_i f^h(\theta_1^h(\omega),\dots,\theta_n^h(\omega))\langle h_i, g_j\rangle g_j \\ &= \sum_{i=1}^{n} \nabla_i f^h(\theta_1^h(\omega),\dots,\theta_n^h(\omega))h_i.\end{aligned}$$

If we only assume that F has finite representation w.r.t. $(h_k)_{k\in\mathbb{N}}$, then there is no reason to assume that its representation w.r.t. another ONB $(g_k)_{k\in\mathbb{N}}$ will also be finite. Instead, we have the following result.

Lemma 3.3.3. *Let $(g_k)_{k\in\mathbb{N}}$ and $(h_k)_{k\in\mathbb{N}}$ be two different orthonormal bases of $L^2(I)$. Let $F \in \mathcal{S}$ with representation $F(\omega) = f^h(\theta_1^h(\omega),\dots,\theta_n^h(\omega))$ for $\omega \in \Omega$ fixed. Then one can find $f^g \in C_b^\infty(\mathbb{R}^\mathbb{N};\mathbb{R})$ such that F has the representation $F(\omega) = f^g((\theta_k^g(\omega))_{k\in\mathbb{N}})$.*

Proof. As $h_i \in L^2(I)$ and $(g_j)_{j\in\mathbb{N}}$ is an ONB of $L^2(I)$, we can write $h_i = \sum_{j\in\mathbb{N}}\langle h_i, g_j\rangle g_j$. Hence

$$\theta_i^h(\omega) = \int_I h_i(s)\mathrm{d}\omega(s) = \int_I \sum_{j\in\mathbb{N}}\langle h_i, g_j\rangle g_j(s)\mathrm{d}\omega(s) = \sum_{j\in\mathbb{N}}\langle h_i, g_j\rangle \theta_j^g(\omega),$$

where we exchange summation and integration, because by Bessel's inequality we have $\sum_{j\in\mathbb{N}} |\langle h_i, g_j\rangle|^2 \le \|h_i\|^2_{L^2(I)} = 1$, where the last equality is due to $(h_i)_{i\in\mathbb{N}}$ being an ONB of $L^2(I)$. Thus f^g is given by

$$f^g\left((\theta_j^g)_{j\in\mathbb{N}}\right) = f^h\left(\sum_{j\in\mathbb{N}}\langle h_1, g_j\rangle\theta_j^g,\dots,\sum_{j\in\mathbb{N}}\langle h_n, g_j\rangle\theta_j^g\right),$$

i.e., $f^g = f^h \circ M$ if we denote by M the $n\times\infty$ matrix with the entries $M_{ij} = \langle h_i, g_j\rangle$. $\square$

3.3.4 Connection to the Malliavin Derivative

Functions $F \in \mathcal{S}$ with representation $F(\omega) = f(\theta_1(\omega),\dots,\theta_n(\omega))$ have the Malliavin derivative $DF(t,\omega) = \sum_{k=1}^{n} \nabla_k f\,(\theta_1(\omega),\dots,\theta_n(\omega))\,h_k(t)$, see for example Definition 4.9 in [Øk97] or Definition 1.2.1 in [Nua06] or cf. Appendix A.2. From the above we can see that the directional derivative can be rewritten as

$$D_\gamma F(\omega) = \int_I DF(t,\omega)\mathrm{d}\gamma(t) = \langle DF(\cdot,\omega),\dot\gamma\rangle_{L^2(I)}. \tag{3.3.3}$$

The same representation holds for $F \in \mathcal{D}$ if and only if the infinite summation and integration can be exchanged. Øksendal [Øk97] introduces two different spaces of Malliavin differentiable functions, $\mathbb{D}_{1,2}$ and $\mathcal{D}_{1,2}$, where the latter is defined roughly as those functions whose directional derivatives have the representation as an integral (3.3.3) for all directions $\gamma \in \mathcal{H}(I)$.

Theorem 3.3.4. *For $F \in \mathcal{D}$, $DF(\cdot, \omega)$ is LS integrable against $\gamma \in \mathcal{H}(I)$ on I if and only if $t \mapsto DF(t, \omega)\dot{\gamma}(t)$ belongs to $L^1(I)$. Sufficient for this is $DF(\cdot, \omega) \in L^2(I)$.*

Proof. It suffices to recall that by definition, if $\gamma \in \mathcal{H}$, then $\dot{\gamma} \in L^2(I)$. The claim then follows with Hölder's inequality. □

Theorem 3.3.5. *If $F \in \mathcal{D}$ such that for a given $\omega \in \Omega$, $DF(\cdot, \omega) \in L^2(I)$ and $\psi(t, \omega) := \sum_{k=1}^{\infty} |\nabla_k f((\theta_j(\omega))_{j\in\mathbb{N}}) h_k(t)\dot{\gamma}(t)| \in L^1(I)$, then $D_\gamma F(\omega) = \int_I DF(t, \omega)\mathrm{d}\gamma(t)$ for $\gamma \in \mathcal{H}$.*

Proof. Fix $\omega \in \Omega$. From Theorem 3.3.4 we know that under the assumptions made, $DF(t, \omega)\dot{\gamma}(t) \in L^1(I)$. For $n \in \mathbb{N}$ let $\phi_n(t, \omega) := \sum_{k=1}^{n} \nabla_k f((\theta_j(\omega))_{j\in\mathbb{N}}) h_k(t)\dot{\gamma}(t)$. Then $(\phi_n)_{n\in\mathbb{N}}$ is dominated by ψ, which allows us to apply Lebesgue's theorem of dominated convergence and thus obtain the desired result. □

It is clear that by allowing more general directions than only Cameron-Martin functions, the space of functions possessing the directional derivatives must become smaller. Our main focus is therefore not on extending Malliavin differentiability, but on defining the directional derivative and finding sufficient conditions for the first equality of (3.3.3) to hold.

3.4 Varying paths with Hölder continuous functions

In this section we extend the previous analysis to directional derivatives for directions γ that are assumed to be α-Hölder continuous functions (we write $\gamma \in C^\alpha(I)$) for some $\alpha > 0$. This is a true extension of directions in the Cameron-Martin space $\mathcal{H}(I)$ because one has the following result:

Lemma 3.4.1. *If $f \in \mathcal{H}(I)$, then $f \in C^{1/2}(I)$.*

Proof. Without loss of generality assume $x > y$ for $x, y \in I$. Then

$$\begin{aligned} |f(x) - f(y)| &= \left| \int_0^x \dot{f}(t)\mathrm{d}t - \int_0^y \dot{f}(t)\mathrm{d}t \right| \\ &\leq \left(\int_y^x |\dot{f}(t)|^2 \mathrm{d}t \right)^{1/2} \cdot \left(\int_y^x 1 \mathrm{d}t \right)^{1/2} \quad \text{(Cauchy-Schwarz inequality)} \\ &\leq \left\| \dot{f} \right\|_{L^2(I)} \cdot |x - y|^{1/2} \end{aligned}$$

and $\left\| \dot{f} \right\|_{L^2(I)} < \infty$ because $f(x) = \int_0^x \dot{f}(t)\mathrm{d}t$ with $\dot{f} \in L^2(I)$. □

We will make extensive use of Ciesielski's isomorphism [Cie60], which we recall here for completeness of the exposition.

Remark 3.4.2. *In this section we want to work with integrals of the structure $\int_I f(t)\mathrm{d}g(t)$ where $g \in C^\alpha(I)$. Let $\widetilde{g}\colon I \to \mathbb{R}$ be defined as $\widetilde{g}(t) := g(t) - g(0)$. Let*

$$C_0^\alpha(I) := \{ \gamma \in C^\alpha(I) \mid \gamma(0) = 0 \} .$$

Then we obviously have $\widetilde{g} \in C_0^\alpha(I)$ and, more importantly, $\int_I f(t)\mathrm{d}g(t) = \int_I f(t)\mathrm{d}\widetilde{g}(t)$. Hence in what fallows it is without loss of generality that we restrict our analysis to integrators belonging to $C_0^\alpha(I)$.

3.4.1 Haar-Schauder basis and Ciesielski's isomorphism

As we have seen in Section 3.3.3, up to a certain degree our results do not depend on the basis chosen for the representation of functions in $\mathcal{S}$ resp. $\mathcal{D}$. Henceforth, we will fix the Haar functions as basis of $L^2(I)$, which are defined as follows (originally introduced by Haar in his dissertation, cf. [Haa10]):

Definition 3.4.3 (Haar functions)**.** *Let $H_{00}(t) \equiv 1$ for $t \in I$ and for $p \in \mathbb{N}_0$ and $1 \leq m \leq 2^p$ define the points*

$$t^0_{pm} := \frac{m-1}{2^p}, \quad t^1_{pm} := \frac{2m-1}{2^{p+1}}, \quad t^2_{pm} := \frac{m}{2^p}$$

and the functions on I given by

$$H_{pm}(t) := \begin{cases} \sqrt{2^p} & : t^0_{pm} < t < t^1_{pm}, \\ -\sqrt{2^p} & : t^1_{pm} < t < t^2_{pm}, \\ \frac{\sqrt{2^p}}{2} & : t = t^0_{pm} \neq 0, \\ -\frac{\sqrt{2^p}}{2} & : t = t^2_{pm} \neq 1, \\ \sqrt{2^p} & : t = t^0_{pm} = 0, \\ -\sqrt{2^p} & : t = t^2_{pm} = 1, \\ 0 & : \textit{else}. \end{cases}$$

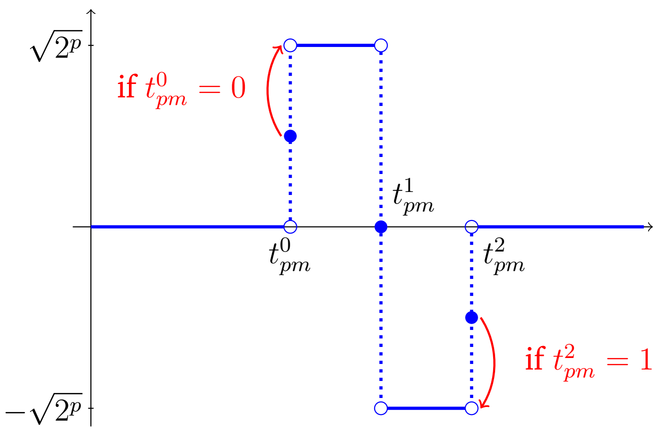


Figure 3.1: Haar function for $p \in \mathbb{N}_0, m \in \{1, \ldots, 2^p\}$

The shape of the Haar functions with the precise jumps can be seen in Figure 3.1.
As long as one considers them as L^2-functions, values at the endpoints of intervals do not matter, hence one can replace that one by the somewhat handier definition that appears for instance in [GIP16]:

Definition 3.4.4 (Haar functions – a version)**.** *For $p \in \mathbb{N}_0$ and $1 \leq m \leq 2^p$, $t \in I$ define*

$$\widetilde{H}_{pm}(t) = \sqrt{2^p} \left[\mathbb{1}_{[t^0_{pm}, t^1_{pm})}(t) - \mathbb{1}_{[t^1_{pm}, t^2_{pm})}(t) \right], \text{ where}$$

$$t^0_{pm} = \frac{m-1}{2^p}, \quad t^1_{pm} = \frac{2m-1}{2^{p+1}}, \quad t^2_{pm} = \frac{m}{2^p}$$

and further let $\widetilde{H}_{00}(t) \equiv 1$ for $t \in I$.

Notation 3.4.5. *Let* $\Lambda := \{(p,m) \mid p \in \mathbb{N}_0, m \in \{1,\dots,2^p\}\}$ *and* $\Lambda_0 := \{(0,0)\} \cup \Lambda$. *With this we introduce for a sequence* $(A_{pm})_{(p,m)\in\Lambda_0}$ *the series*

$$\sum_{(p,m)\in\Lambda} A_{pm} := \sum_{p=0}^{\infty}\sum_{m=1}^{2^p} A_{pm} \quad \text{and} \quad \sum_{(p,m)\in\Lambda_0} A_{pm} := A_{00} + \sum_{p=0}^{\infty}\sum_{m=1}^{2^p} A_{pm}.$$

Remark 3.4.6 (Different versions of the definition). *There exist different versions of definitions of Haar functions in the literature, two of which have been presented above. Alfred Haar introduced his functions originally in such a way that the Fourier series associated to a continuous function converges uniformly to that function*[6]*, i.e., for* $f \in C(I)$ *and coefficients* $a_{pm} := \int_I f(t)H_{pm}(t)\mathrm{d}t$ *for* $(p,m) \in \Lambda_0$*, one has uniform convergence of* $\sum_{(p,m)\in\Lambda_0} a_{pm}H_{pm}$ *to* f. *It was remarked in* [Ul'64] *that the definition of Haar functions given in* [KS51] *results in a loss of this property.* [KS51] *defined Haar functions as follows:*

$$H^{KS}_{pm}(t) = \sqrt{2^p}\left[\mathbb{1}_{[t^0_{pm},t^1_{pm})}(t) - \mathbb{1}_{(t^1_{pm},t^2_{pm}]}(t)\right] \quad \text{and} \quad H^{KS}_{00} \equiv 1. \tag{3.4.1}$$

In [Ul'64, page 3] *the following counterexample shows that the slightly changed definition results in a loss of the uniform convergence property stated by Haar. One may evaluate the Fourier series of the function*

$$f(t) = \begin{cases} 1 & : 0 \le t \le \frac{3}{4}, \\ 25 - 32t & : \frac{3}{4} < t \le 1 \end{cases}$$

at $t = \frac{1}{2}$*, which yields* 2*, while* $f(\frac{1}{2}) = 1$. *The same problem arises with Definition 3.4.4 when evaluating* f *and its Fourier series at* $t = 1$. *In fact, any function* $f \in C(I)$ *with* $f(1) \ne \int_I f(t)\mathrm{d}t$ *is a valid counterexample for showing that the Haar functions from Definition 3.4.4 don't have the uniform convergence property. This highlights that whenever results are based on a pointwise evaluation of Haar functions, endpoints do matter.*

The following functions, which go back to Faber [Fab10] and Schauder[7] [Sch27], allow us to write any continuous function as linear combination of basis functions. While Haar functions are discontinuous and finite linear combinations thereof cannot yield a continuous function, Faber-Schauder functions are already continuous. The following definition is independent of the chosen definition of the Haar functions, i.e., the choice of the values at the endpoints is irrelevant.

Definition 3.4.7 (Faber-Schauder functions on I). *For* $t \in I$ *define* $G_{00}(t) := t$ *and*

$$G_{pm}(t) := \int_0^t H_{pm}(s)\mathrm{d}s = \begin{cases} \sqrt{2^p}(t - t^0_{pm}), & t \in [t^0_{pm}, t^1_{pm}) \\ \sqrt{2^p}(t^2_{pm} - t), & t \in [t^1_{pm}, t^2_{pm}) \\ 0, & \text{otherwise} \end{cases} \qquad \text{for } (p,m) \in \Lambda.$$

Following [Cie59] and [GIP16] we shall call these functions simply Schauder functions. For any continuous function $f \in C(I)$ write

$$f_{pm} := \langle H_{pm}, \mathrm{d}f\rangle := \sqrt{2^p}\left[2f(t^1_{pm}) - f(t^0_{pm}) - f(t^2_{pm})\right], \quad \text{for } (p,m) \in \Lambda$$

[6] This result is cited in the Appendix as Theorem B.2.1 for later reference.

[7] Schauder describes his functions with a different normalization s.t. the functions' highest value is 1 for any $(p,m) \in \Lambda_0$.

and additionally $f_{00} := \langle H_{00}, \mathrm{d}f\rangle = f(1) - f(0)$. For a more convenient presentation of Ciesielski's isomorphism, we furthermore introduce the rescaled functions

$$\xi_{pm}^{(\alpha)}(t) := \frac{2^{(p+1)\alpha}}{2\sqrt{2^p}} H_{pm}(t), \quad \phi_{pm}^{(\alpha)}(t) := \frac{2\sqrt{2^p}}{2^{(p+1)\alpha}} G_{pm}(t), \qquad \text{for } (p,m) \in \Lambda,$$

and

$$\xi_{00}^{(\alpha)}(t) \equiv 1 \quad , \text{ and } \quad \phi_{00}^{(\alpha)}(t) := t$$

for $t \in I$ and $\alpha \in (0,1)$. The sequence $(\phi_{pm}^{(\alpha)})_{(p,m)\in\Lambda_0}$ is normalized such that $\left|\phi_{pm}^{(\alpha)}\right|_\alpha = 1$. With this we recall from [Cie60][8]:

Theorem 3.4.8. *For $\alpha \in (0,1)$ the following is an isomorphism:*

$$T_\alpha \colon \begin{cases} (\ell^\infty, \|\cdot\|_{\ell^\infty}) & \longrightarrow (C_0^\alpha(I), |\cdot|_\alpha) \\ (\xi_{pm})_{(p,m)\in\Lambda_0} & \longmapsto x := \sum_{(p,m)\in\Lambda_0} \xi_{pm}\phi_{pm}^{(\alpha)}, \end{cases}$$

$$T_\alpha^{-1} \colon \begin{cases} (C_0^\alpha(I), |\cdot|_\alpha) & \longrightarrow (\ell^\infty, \|\cdot\|_{\ell^\infty}) \\ x & \longmapsto \left(\xi_{pm} := \int_0^1 \xi_{pm}^{(\alpha)}(s)\mathrm{d}x(s)\right)_{(p,m)\in\Lambda_0}. \end{cases}$$

Moreover, the operator norms satisfy

$$\frac{2}{3(2^\alpha-1)(2^{1-\alpha}-1)} \le \|T_\alpha\| \le \frac{2}{(2^\alpha-1)(2^{1-\alpha}-1)}, \qquad \left\|T_\alpha^{-1}\right\| = 1. \tag{3.4.2}$$

Note that we can write $\gamma \in C_0^\alpha(I)$ in two ways:

$$\gamma = \sum_{(p,m)\in\Lambda_0} \widetilde{\gamma}_{pm}^{(\alpha)}\phi_{pm}^{(\alpha)} = \sum_{(p,m)\in\Lambda_0} \gamma_{pm}G_{pm}, \tag{3.4.3}$$

where for $(p,m) \in \Lambda_0$

$$\widetilde{\gamma}_{pm}^{(\alpha)} = c_{pm}^{(\alpha)} \cdot \gamma_{pm} \quad \text{ if we let } \quad c_{pm}^{(\alpha)} := \begin{cases} 2^{p(\alpha-\frac{1}{2})+\alpha-1} & ,(p,m)\in\Lambda, \\ 1 & ,(p,m) = (0,0). \end{cases} \tag{3.4.4}$$

What exactly do we know about $(\gamma_{pm})_{(p,m)\in\Lambda_0}$?
Observe that $c_{pm}^{(\alpha)} > 0$ for all $(p,m) \in \Lambda_0$ and $\alpha \in (0,1)$ and that $\frac{\partial c_{pm}^{(\alpha)}}{\partial \alpha} = (p+1)\cdot \ln 2 \cdot c_{pm}^{(\alpha)}$. Furthermore, $\frac{\partial c_{pm}^{(\alpha)}}{\partial p} = (\alpha - \frac{1}{2}) \cdot \ln 2 \cdot c_{pm}^{(\alpha)}$, hence

$$\frac{\partial c_{pm}^{(\alpha)}}{\partial p} \begin{cases} < 0 & \text{if } \alpha < \frac{1}{2}, \\ = 0 & \text{if } \alpha = \frac{1}{2}, \\ > 0 & \text{if } \alpha > \frac{1}{2}. \end{cases}$$

This observation allows us to formulate the following result.

Lemma 3.4.9. *With the above notation, in particular* (3.4.3) *and* (3.4.4)*, we have*

[8] This result does not require a pointwise evaluation of Haar functions.

(a) $(\widetilde{\gamma}_{pm}^{(1/2)})_{(p,m)\in\Lambda_0} \in \ell^\infty \iff (\gamma_{pm})_{(p,m)\in\Lambda_0} \in \ell^\infty$;

(b) for $\alpha \in \left(0, \frac{1}{2}\right)$, $(\gamma_{pm})_{(p,m)\in\Lambda_0} \in \ell^\infty$ *implies* $(\widetilde{\gamma}_{pm}^{(\alpha)})_{(p,m)\in\Lambda_0} \in \ell^\infty$;

(c) for $\alpha \in \left(\frac{1}{2}, 1\right)$, $(\widetilde{\gamma}_{pm}^{(\alpha)})_{(p,m)\in\Lambda_0} \in \ell^\infty$ *implies* $(\gamma_{pm})_{(p,m)\in\Lambda_0} \in \ell^\infty$.

We can also directly state two further results.

Lemma 3.4.10 ([Imk15, Lemma 1.1]). *A continuous function* $f\colon I \to \mathbb{R}$ *lies in* $C_0^\alpha(I)$ *if and only if* $\sup_{(p,m)\in\Lambda_0} 2^{p(\alpha-\frac{1}{2})}\,|f_{pm}| < \infty$, *i.e., if* $\left(2^{p(\alpha-\frac{1}{2})} f_{pm}\right)_{(p,m)\in\Lambda_0} \in \ell^\infty$.

Lemma 3.4.11. *Let* $\alpha \in (0, \frac{1}{2})$ *and* $\gamma \in C_0^\alpha(I)$. *Let* $(\beta_{pm})_{(p,m)\in\Lambda_0}$ *be a sequence with values in* $\mathbb{R}^1$. *Further assume that*

$$\left(\frac{\beta_{pm}}{c_{pm}^{(\alpha)}}\right) \in \ell^1.$$

Then $\sum_{(p,m)\in\Lambda_0} \beta_{pm}\gamma_{pm}$ *converges unconditionally*[9] *(in* $\mathbb{R}^1$*).*

Proof. By [Cie60], $\gamma \in C^\alpha$ implies that $\widetilde{\gamma}_{pm}^{(\alpha)} = c_{pm}^{(\alpha)}\gamma_{pm}$ is a bounded sequence. Apply Lemma B.2.6 to attain the result. □

We conclude this subsection by presenting an example that demonstrates the applicability of Ciesielski's isomorphism.

Example 3.4.12. *For any* $\alpha \in (0,1)$, $f(t) := t^\alpha$ *is in* $C^\beta(I)$ *for any* $0 < \beta \le \alpha$, *but not for any* $\beta > \alpha$. *Observe that for* $\alpha < \frac{1}{2}$, $\gamma \notin \mathcal{H}$, *because* $\dot\gamma \notin L^2(I)$:

$$\|\dot\gamma\|_{L^2(I)}^2 = \int_0^1 (\dot\gamma(t))^2 \mathrm{d}t = \int_0^1 \left(\alpha t^{\alpha-1}\right)^2 \mathrm{d}t = \begin{cases} \frac{\alpha^2}{2\alpha-1}, & \alpha > \frac{1}{2} \\ \infty, & \alpha \le \frac{1}{2}. \end{cases}$$

By Ciesielski, $f(t) = \sum_{(p,m)\in\Lambda_0} f_{pm}G_{pm}(t)$ *for* $f_{pm} = \int_0^1 H_{pm}(t)\mathrm{d}f(t)$. *In this particular case we calculate that for* $(p,m) \in \Lambda$

$$\begin{aligned} f_{pm} &= \sqrt{2^p}\left[2f(t_{pm}^1) - f(t_{pm}^0) - f(t_{pm}^2)\right] \\ &= 2^{p/2-(p+1)\alpha}\left[2(2m-1)^\alpha - (2m-2)^\alpha - (2m)^\alpha\right] \\ &=: 2^{p/2-(p+1)\alpha} M(m,\alpha). \end{aligned}$$

It can be shown that $f_{pm} \ge 0$ *for* $(p,m) \in \Lambda_0$ *and that it decreases as* p *and/or* m *increase. For more details, see Lemma B.3.1.*

In the notation from Ciesielski's isomorphism, $f = \sum_{(p,m)\in\Lambda_0} \xi_{pm}\phi_{pm}^{(\beta)}$, *where the coefficients* $(\xi_{pm})_{(p,m)\in\Lambda_0}$ *are in* ℓ^∞ *if and only if* $f \in C_0^\beta(I)$. *They can then be calculated as*

$$\xi_{pm} = \frac{2^{(p+1)\beta}}{2\sqrt{2^p}} \underbrace{\int_0^1 H_{pm}(t)\mathrm{d}f(t)}_{f_{pm}} = \frac{2^{(p+1)\beta}}{2\sqrt{2^p}} \cdot \frac{\sqrt{2^p}}{2^{(p+1)\alpha}} M(m,\alpha) = 2^{(p+1)(\beta-\alpha)} \cdot \frac{1}{2} M(m,\alpha).$$

Now $\frac{1}{2}M(m,\alpha)$ *is bounded from above by e.g.* $M(1,\alpha) = 2 - 2^\alpha$. *The factor with* p *is either bounded by 1 if* $\beta \le \alpha$ *or unbounded if* $\beta > \alpha$, *hence indeed the sequence* $(\xi_{pm})_{(p,m)\in\Lambda_0}$ *is in* ℓ^∞ *if and only if* $\beta \le \alpha$, *which is exactly the Hölder regularity of* $f(t) = t^\alpha$ *on* $I = [0,1]$.

[9] See Definition B.2.5 in the Appendix.

3.4.2 Analysis of the directional derivative in Hölder directions

Throughout the rest of this section we assume that $\gamma \in C_0^\alpha(I)$ with $\alpha \in (0,1)$ with Schauder expansion $\gamma(t) = \sum_{(p,m)\in\Lambda_0} \gamma_{pm} G_{pm}(t)$ for $t \in I$. Ciesielski's isomorphism will help us by providing properties of $(\gamma_{pm})_{(p,m)\in\Lambda_0}$ depending on the parameter α. As in the previous section we start again with cylinder functions for which we define the derivative in direction γ and then extend this to functions with infinite representation, i.e. $F \in \mathcal{D}$.

Notation 3.4.13. *For $N \in \mathbb{N}$ write $\Lambda^N := \{(p,m) \mid p \in \{0,1,\dots,N\}, m \in \{1,\dots,2^p\}\}$ and $\Lambda_0^N := \{(0,0)\} \cup \Lambda^N$.*

With the above notation and $\theta_{pm}(\omega) := \int_I H_{pm}(t)\mathrm{d}\omega(t)$ for $(p,m) \in \Lambda_0$, let us consider $F \in \mathcal{S}$ with representation $F(\omega) = f\left((\theta_{pm}(\omega))_{(p,m)\in\Lambda_0^N}\right)$.
We want to write, starting with the same arguments as in (3.3.1),

$$\begin{aligned} D_\gamma F(\omega) &:= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left[F(\omega+\varepsilon\gamma) - F(\omega)\right] \\ &= \sum_{(p,m)\in\Lambda_0^N} \nabla_{pm} f\left((\theta_{pm}(\omega))_{(p,m)\in\Lambda_0^N}\right)\langle H_{pm}, \mathrm{d}\gamma\rangle \\ &= \sum_{(p,m)\in\Lambda_0^N} \nabla_{pm} f\left((\theta_{pm}(\omega))_{(p,m)\in\Lambda_0^N}\right)\gamma_{pm}, \end{aligned} \tag{3.4.5}$$

where the first equality holds due to the assumptions on f. To see that the second equality holds, recall that [Cie59, Theorem 3] yields the following:

Proposition 3.4.14. *If $\gamma = \sum_{(q,n)\in\Lambda_0^N} \gamma_{qn} G_{qn}$, then the coefficients γ_{pm} are explicitly given by $\gamma_{pm} = \langle H_{pm}, \mathrm{d}\gamma\rangle$ for $(p,m) \in \Lambda_0^N$.*

As before, we can write $D_\gamma F(\omega) = \langle DF(\cdot,\omega), \mathrm{d}\gamma\rangle$ with Malliavin derivative

$$DF(t,\omega) = \sum_{(p,m)\in\Lambda_0^N} \nabla_{pm} f\left((\theta_{qn}(\omega))_{(q,n)\in\Lambda_0^N}\right) H_{pm}(t).$$

In what follows, the Malliavin derivative of $F \in \mathcal{D}$ is given by

$$DF(t,\omega) = \sum_{(p,m)\in\Lambda_0} \nabla_{pm} f\left((\theta_{qn}(\omega))_{(q,n)\in\Lambda_0}\right) H_{pm}(t),$$

provided the series converges for fixed $\omega \in \Omega$ for all $t \in I$.

3.4.3 Variation of functions in $\mathcal{D}$ in Hölder directions

Fix $\omega \in \Omega$ and let $F \in \mathcal{D}$ with representation $F(\omega) = f\left((\theta_{pm}(\omega))_{(p,m)\in\Lambda_0}\right)$. For brevity of notation, let $\beta_{pm}(\omega) := \nabla_{pm} f\left((\theta_{qn}(\omega))_{(q,n)\in\Lambda_0}\right)$ for $(p,m) \in \Lambda_0$. As before, we are interested in the directional derivative, which, if it exists, has the structure

$$
\begin{aligned}
D_\gamma F(\omega) &:= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left[F(\omega+\varepsilon\gamma)-F(\omega)\right]\\
&= \lim_{N\to\infty}\sum_{(p,m)\in\Lambda_0^N}\beta_{pm}(\omega)\int_I H_{pm}(t)\mathrm{d}\gamma(t)\\
&= \lim_{N\to\infty}\sum_{(p,m)\in\Lambda_0^N}\beta_{pm}(\omega)\gamma_{pm},
\end{aligned}
\tag{3.4.6}
$$

and may or may not be equal to the expression

$$
\langle DF(\cdot,\omega),\mathrm{d}\gamma\rangle = \int_I DF(t,\omega)\mathrm{d}\gamma(t) = \int_I \sum_{(p,m)\in\Lambda_0}\beta_{pm}(\omega)H_{pm}(t)\mathrm{d}\gamma(t).
$$

Our main goal is to establish the existence of $D_\gamma F$, which requires the existence of $\int_I H_{pm}(t)\mathrm{d}\gamma(t)$ and the convergence of the series (3.4.6), which is the candidate for $D_\gamma F$. In addition, we also ask whether, provided $DF(t,\omega)$ exists for all $t\in I$, the integral $\int_I DF(t,\omega)\mathrm{d}\gamma(t)$ exists and whether it equals $D_\gamma F(\omega)$. We proceed as follows:

Step 1: Find conditions for the convergence of $\sum_{(p,m)\in\Lambda_0^N}\beta_{pm}(\omega)\gamma_{pm}$ as $N\to\infty$.

Step 2: Find conditions such that the integral $\int_I DF(t,\omega)\mathrm{d}\gamma(t)$ is well defined provided $DF(t,\omega)=\sum_{(p,m)\in\Lambda_0}\beta_{pm}(\omega)H_{pm}(t)$.

Step 3: Find conditions such that

$$
\int_I \lim_{N\to\infty}\sum_{(p,m)\in\Lambda_0^N}\beta_{pm}H_{pm}(t)\mathrm{d}\gamma(t) = \lim_{N\to\infty}\int_I\sum_{(p,m)\in\Lambda_0^N}\beta_{pm}H_{pm}(t)\mathrm{d}\gamma(t).
$$

Step 1 – The directional derivative

Theorem 3.4.15. *Fix $\omega\in\Omega$ and let $F(\omega)=f((\theta_{pm}(\omega))_{(p,m)\in\Lambda_0})\in\mathcal{D}$ with $\beta_{pm}(\omega):=\nabla_{pm}f((\theta_{qn}(\omega))_{(q,n)\in\Lambda_0})$ for $(p,m)\in\Lambda_0$. For some $\alpha\in(0,1)$ let $\gamma\in C_0^\alpha(I)$ with Schauder representation $\gamma(t)=\sum_{(p,m)\in\Lambda_0}\gamma_{pm}G_{pm}(t)$. Assume that*

$$
\left(2^{p(\frac{1}{2}-\alpha)}\beta_{pm}(\omega)\right)_{(p,m)\in\Lambda_0}\in\ell^1.
$$

Then $\sum_{(p,m)\in\Lambda_0}\beta_{pm}(\omega)\gamma_{pm}$ converges absolutely and its limit is $D_\gamma F(\omega)$.

Proof. Theorem 3.4.8 tells us that for $\gamma\in C_0^\alpha(I)$ with the given Schauder representation, $\gamma_{pm}^{(\alpha)}:=\frac{2^{(p+1)\alpha}}{2\sqrt{2^p}}\gamma_{pm}\in\ell^\infty$ where $\gamma_{pm}=\int_0^1 H_{pm}(t)\mathrm{d}\gamma(t)$. Thus,

$$
\sum_{(p,m)\in\Lambda_0}\beta_{pm}(\omega)\gamma_{pm} = \sum_{(p,m)\in\Lambda_0}\beta_{pm}(\omega)\frac{2^{1+\frac{p}{2}}}{2^{(p+1)\alpha}}\gamma_{pm}^{(\alpha)} = 2^{1-\alpha}\sum_{(p,m)\in\Lambda_0}2^{p(\frac{1}{2}-\alpha)}\beta_{pm}(\omega)\gamma_{pm}^{(\alpha)}.
$$

With Lemma B.2.7 and the assumptions made, the series converges absolutely. □

The second point in the proof allows us to find a condition that guarantees that F possesses directional derivatives in ω for α-Hölder continuous directions for any $\alpha>0$.

Corollary 3.4.16. *Let $\omega \in \Omega$ be such that $(2^{p/2}\beta_{pm}(\omega))_{(p,m)\in\Lambda_0} \in \ell^1$. Then $D_\gamma F(\omega)$ exists for all $\gamma \in C^\alpha(I)$, where α can be arbitrarily chosen in $(0,1)$.*

Proof. This follows directly from Theorem 3.4.15, because $(2^{-p\alpha})_{p\in\mathbb{N}}$ is a null sequence for any $\alpha \in (0,1)$. □

Let us briefly explain the difference between the above condition on f and the requirements on f for belonging to a Besov space.

Besov spaces appear in the literature with different definitions depending on the context. In [Roy93] one can find the representation for $\alpha > 1/p$ given by

$$\mathcal{B}^\alpha_{p,q} := \left\{ f\colon [0,1] \to \mathbb{R} \,\middle|\, \|f\|_{\alpha,p,q} < \infty \right\} \text{ where}$$

$$\|f\|_{\alpha,p,q} \text{ is equivalent to } \left(\sum_j 2^{-jq(1/2-\alpha+1/p)} \left(\sum_{k=1}^{2^j} |f_{jk}|^p \right)^{q/p} \right)^{1/q}$$

if $f_{jk} = \int_I H_{jk}(t)\mathrm{d}f(t) = 2^{j/2}\left[2f(t^1_{jk}) - f(t^0_{jk}) - f(t^2_{jk})\right]$. One can observe that while Besov spaces are concerned with the sequence given by the coefficients from the Haar-Schauder representation ($f = \sum_{(p,m)\in\Lambda_0} f_{pm}G_{pm}$), we search for properties of the sequence given by the (infinitely many) partial derivatives of f.

Step 2 – Integrability of DF

The result we want to show is that if $\gamma \in C^\alpha_0(I)$ and if DF has finite q-variation for $q < \frac{1}{1-\alpha}$, then $\int_I DF(t,\omega)\mathrm{d}\gamma(t)$ is a well-defined integral in Young's setting.[10] To make this rigorous, we first recall details from the Young integral:

In [You36] the conditions for the existence of the RS integral $\int_I f(t)\mathrm{d}g(t)$ are extended from g being continuous and f of bounded variation[11] to both being of finite p- and q-variation, respectively, with $\frac{1}{p} + \frac{1}{q} > 1$ and having no common discontinuities. In particular neither f nor g is required to be continuous. In later works (e.g. [GIP16]) slightly stricter requirements were made, i.e., f and g are supposed to be α- and β-Hölder continuous, respectively, for $\alpha + \beta > 1$. Our integrand $DF(t,\omega) = \sum \beta_{pm}(\omega)H_{pm}(t)$ is by nature not necessarily continuous, but instead for $F \in \mathcal{S}$ it has a finite and for $F \in \mathcal{D}$ at most a countable number of jumps. If DF has finite p-variation and if γ is α-Hölder-continuous, we can assure that integrand and integrator have no common discontinuities and that the (Young)-Stieltjes integral exists, provided $\alpha + \frac{1}{p} > 1$.

For a link between the spaces $C^\alpha(I)$ and $\mathcal{V}_p(I)$ we refer to Theorem B.2.12 and its discussion. More results on p-variation and the corresponding spaces can be found in Appendix B.2.4.

Example 3.4.17. *Let $\alpha, \beta > 0$ and define functions $f(t) = t^\alpha$ and $g(t) = t^\beta$ for $t \in I$, which are in $C^\alpha_0(I)$ and $C^\beta_0(I)$, respectively. Then $\int_I f(t)\mathrm{d}g(t) = \int_I f(t)\dot{g}(t)\mathrm{d}t = \frac{\beta}{\alpha+\beta}$ exists even if $\alpha + \beta$ is not strictly bigger than 1. This is not surprising, as $f, g \in BV(I) \cap C_0(I)$*

[10] Under these conditions, $\alpha + \frac{1}{q} > 1$ and as long as integrand or integrator is continuous, there is no problem with points of discontinuity.

[11] We use the terms *finite (p-)variation* and *bounded (p-)variation* synonymously.

and it is known that if one function is continuous and the other of bounded variation, the corresponding (Riemann-)Stieltjes integral exists. This example illustrates that the Hölder condition is indeed strictly stronger than p-variation combined with continuity.

Recall that $V_q(f)$ denotes the q-variation of f on our fixed interval $I = [0,1]$. Then by lower semi-continuity of the q-variation (see Lemma B.2.11),

$$V_q(DF) \leq \liminf_{N\to\infty} V_q(DF_N) \quad \text{for} \quad DF_N := \sum_{(p,m)\in\Lambda_0^N} \beta_{pm} H_{pm}.$$

First we derive conditions which guarantee that DF has finite q-variation. To this end, observe that each summand $\beta_{pm}H_{pm}$ adds at most $(2+2^q)\cdot 2^{\frac{pq}{2}}\left|\beta_{pm}^q\right|$ to the q-variation, hence

$$V_q(DF) \leq \sum_{(p,m)\in\Lambda_0} (2+2^q)\cdot 2^{\frac{pq}{2}}\left|\beta_{pm}^q\right|.$$

If the above expression is finite, DF has finite q-variation. A sufficient condition is that the sequence $\left(2^{p/2}\beta_{pm}\right)_{(p,m)\in\Lambda_0} \in \ell^q$, which is a weaker condition than for that sequence to be in ℓ^1.

Theorem 3.4.18. *Let $\alpha \in (0,1)$. Fix $\omega \in \Omega$, take $F(\omega) = f((\theta_{pm}(\omega))_{(p,m)\in\Lambda_0})$ and set $\beta_{pm}(\omega) := \nabla_{pm} f((\theta_{qn}(\omega))_{(q,n)\in\Lambda_0})$. Further, let $\gamma \in C_0^\alpha(I)$ with Schauder representation $\gamma(t) = \sum_{(p,m)\in\Lambda_0}\gamma_{pm}G_{pm}(t)$. Assume that the sequence $(2^{p/2}|\beta_{pm}|)_{(p,m)\in\Lambda_0} \in \ell^q$ for $q < \frac{1}{1-\alpha}$. Then $DF(\cdot,\omega)$ has finite q-variation and $\int_I DF(t,\omega)\mathrm{d}\gamma(t)$ is well defined, i.e., $DF(\cdot,\omega)$ is Stieltjes integrable on I w.r.t. γ.*

Proof. If for given $\omega \in \Omega$, $DF(\cdot,\omega)$ has finite q-variation and $\gamma \in C^\alpha(I)$ for $\alpha + \frac{1}{q} > 1$ (or equivalently $q < \frac{1}{1-\alpha}$), then [You36] (*Theorem on Stieltjes integrability*) states that the integral $\int_I DF(t,\omega)\mathrm{d}\gamma(t)$ is well defined. The calculations preceeding the theorem show that under the assumptions made, indeed $DF(\cdot,\omega) \in \mathcal{V}_q(I)$, hence the proof is complete. □

Step 3 – Is $\langle DF, \mathrm{d}\gamma\rangle = D_\gamma F$?

We work under the assumptions made in the previous steps, i.e., for fixed $\omega \in \Omega$ the sequence of coefficients $(\beta_{pm}(\omega))_{(p,m)\in\Lambda_0}$ associated to F must lie in the set

$$\ell^*(\alpha) := \left\{ \begin{array}{r} (\beta_{pm})_{(p,m)\in\Lambda_0} \in \ell^2 \mid \left(2^{(\frac{1}{2}-\alpha)p}\beta_{pm}\right)_{(p,m)\in\Lambda_0} \in \ell^1 \text{ and} \\ \exists\, q < \dfrac{1}{1-\alpha} \text{ s.t. } \left(2^{\frac{p}{2}}\beta_{pm}\right)_{(p,m)\in\Lambda_0} \in \ell^q \end{array} \right\}. \tag{3.4.7}$$

These conditions guarantee that $D_\gamma F(\omega)$ exists and that $DF(\cdot,\omega)$ is integrable on I w.r.t. γ. Our task is to find sufficient conditions that allow us to exchange the given Stieltjes integral with infinite summation.

Application of dominated convergence

Fix $\gamma \in C_0^\alpha(I)$ and $\omega \in \Omega$ and let $\varphi_N(t,\omega) := \sum_{(p,m)\in\Lambda_0^N}\beta_{pm}(\omega)H_{pm}(t)$. Clearly $\varphi_N \in \mathcal{L}(\gamma)$, because φ_N has only a finite number of jumps and is otherwise constant (w.r.t. t).

φ_N converges to $DF(\cdot,\omega)$ uniformly on I (Corollary B.2.9) and $\varphi_N(\cdot,\omega)$ is uniformly dominated by $\psi(t,\omega) := \sum_{(p,m)\in\Lambda_0} |\beta_{pm}(\omega)H_{pm}(t)|$. From previous theorems we know that, provided $(\beta_{pm})_{(p,m)\in\Lambda_0}$ lies in the set $\ell^*(\alpha)$ described in (3.4.7), $DF \in \mathcal{V}_q(I)$, hence the inequality $V_q(\psi(\cdot,\omega)) \leq V_q(DF(\cdot,\omega))$ tells us that $\psi \in \mathcal{V}_q(I) \subset \mathcal{R}(\gamma) \subset \mathcal{L}(\gamma)$.

Theorem 3.4.19. *Fix $\omega \in \Omega$. Let $F \in \mathcal{D}$ be represented by f as before. For $\alpha \in (0,1)$ let $\gamma \in C_0^\alpha(I)$. Assume that the sequence $(\beta_{pm}(\omega))_{(p,m)\in\Lambda_0} \in \ell^*(\alpha)$. Then $DF(\cdot,\omega) \in \mathcal{R}(\gamma) \subset \mathcal{L}(\gamma)$ and*

$$\langle DF, \mathrm{d}\gamma\rangle = \int_I \lim_{N\to\infty} \sum_{(p,m)\in\Lambda_0^N} \beta_{pm}H_{pm}(t)\mathrm{d}\gamma(t) = \lim_{N\to\infty}\int_I \sum_{(p,m)\in\Lambda_0^N} \beta_{pm}H_{pm}(t)\mathrm{d}\gamma(t) = D_\gamma F.$$

Example 3.4.20. *Let $\gamma\colon I \to \mathbb{R}$ be given by $\gamma(t) = t^\alpha$ for $\alpha \in (0,1)$. From Example 3.4.12 we know that $\gamma \in C_0^\alpha(I)$.*
Let $F(\omega) = \sum_{p=0}^\infty a_p\theta_{p1}(\omega)$ where the sequence $(a_p)_{p\in\mathbb{N}_0}$ shall be given by $a_p = \frac{1}{2^p}$, $p \in \mathbb{N}_0$. In our previous notation we have

$$\beta_{pm} = \begin{cases} \frac{1}{2^p} & , p\in\mathbb{N}_0, m=1, \\ 0 & , \textit{else}. \end{cases}$$

We want to show that Theorem 3.4.19 can be applied to these data and we explicitly calculate $\langle DF, \mathrm{d}\gamma\rangle$ and $D_\gamma F$.
We clearly have $F \in \mathcal{D}$, hence we know that $DF(t,\omega) = \sum_{(p,m)\in\Lambda_0}\beta_{pm}(\omega)H_{pm}(t)$, which is independent of ω by construction. We further calculate

$$2^{\frac{p}{2}}\beta_{pm} = \begin{cases} 2^{-\frac{p}{2}} & , p\in\mathbb{N}_0, m=1, \\ 0 & , \textit{else}, \end{cases}$$

thus

$$\sum_{(p,m)\in\Lambda_0} \left|2^{\frac{p}{2}}\beta_{pm}\right| = \sum_{p=0}^\infty 2^{-\frac{p}{2}} = \sum_{p=0}^\infty \left(\frac{1}{\sqrt{2}}\right)^p = \frac{1}{1-\frac{1}{\sqrt{2}}} < \infty.$$

We conclude that $(2^{p/2}\beta_{pm})_{(p,m)\in\Lambda_0} \in \ell^1$ and we recall that $\gamma \in C_0^\alpha(I)$; with these results in mind we can apply Theorem 3.4.19. Let us quickly verify that indeed $\langle DF, \mathrm{d}\gamma\rangle = D_\gamma F$ by calculating both terms. By definition we have

$$D_\gamma F = \sum_{(p,m)\in\Lambda_0} \beta_{pm}\int_0^1 H_{pm}(t)\mathrm{d}\gamma(t) = \sum_{p=0}^\infty a_p\gamma_{p1},$$

where

$$\begin{aligned}\gamma_{p1} &= \int_0^1 H_{p1}(t)\mathrm{d}\gamma(t) = \sqrt{2^p}\left[\int_{t_{p1}^0}^{t_{p1}^1}\mathrm{d}\gamma(t) - \int_{t_{p1}^1}^{t_{p1}^2}\mathrm{d}\gamma(t)\right] \\ &= \sqrt{2^p}\left[2\left(\frac{2-1}{2^{p+1}}\right)^\alpha - \left(\frac{2-2}{2^{p+1}}\right)^\alpha - \left(\frac{2-0}{2^{p+1}}\right)^\alpha\right] \\ &= \frac{\sqrt{2^p}}{2^{\alpha(p+1)}}\left[2\cdot 1 - 0 - 2^\alpha\right] = 2^{-\alpha}\left(2-2^\alpha\right)\cdot 2^{\left(\frac{1}{2}-\alpha\right)p}.\end{aligned}$$

With this we can calculate

$$D_\gamma F = \sum_{p=0}^{\infty} 2^{-p} \cdot 2^{-\alpha}\left(2-2^\alpha\right) \cdot 2^{\left(\frac{1}{2}-\alpha\right)p} = 2^{-\alpha}\left(2-2^\alpha\right)\sum_{p=0}^{\infty}\left(\frac{1}{2^{\alpha+\frac{1}{2}}}\right)^p = \frac{2^{\frac{1}{2}}\left(2-2^\alpha\right)}{2^{\alpha+\frac{1}{2}}-1}.$$

For the scalar product, first observe that $DF(1) = -1$ and for $t \in [0,1)$,

$$DF(t) = \sum_{p=0}^{\infty} 2^{-\frac{p}{2}}\left[\mathbb{1}_{\left[0,\frac{1}{2^{p+1}}\right)}(t) - \mathbb{1}_{\left(\frac{1}{2^{p+1}},\frac{2}{2^{p+1}}\right)}(t) - \frac{1}{2}\mathbb{1}_{\left\{\frac{2}{2^{p+1}}\right\}}(t)\right],$$

hence by dividing the interval I into dyadic intervals and elementary calculus we obtain

$$\begin{aligned}
\langle DF, \mathrm{d}\gamma\rangle &= \int_I DF(t)\mathrm{d}\gamma(t) \\
&= \sum_{p=0}^{\infty}\int_{\frac{1}{2^{p+1}}}^{\frac{1}{2^p}}\left(\left(\sum_{q=0}^{p-1}2^{-\frac{q}{2}}\right) - 2^{-\frac{p}{2}}\right)\mathrm{d}\gamma(t) \\
&= \sum_{p=0}^{\infty}\left[\left(\frac{1}{2^\alpha}\right)^p - \left(\frac{1}{2^\alpha}\right)^{p+1}\right]\cdot\left[\frac{1-2^{-\frac{p}{2}}}{1-2^{-\frac{1}{2}}} - \frac{1}{2^{-\frac{p}{2}}}\right] \\
&= -1 + \sum_{p=0}^{\infty}\left(\frac{1}{2^\alpha}\right)^{p+1}\left((-1)\cdot\left[\frac{1-2^{-\frac{p}{2}}}{1-2^{-\frac{1}{2}}} - \frac{1}{2^{-\frac{p}{2}}}\right] + \left[\frac{1-2^{-\frac{p+1}{2}}}{1-2^{-\frac{1}{2}}} - \frac{1}{2^{-\frac{p+1}{2}}}\right]\right) \\
&= -1 + \frac{4-2^{\frac{1}{2}}}{2^{\alpha+1}}\sum_{p=0}^{\infty}\left(\frac{1}{2^{\alpha+\frac{1}{2}}}\right)^p \\
&= \frac{2^{\frac{3}{2}} - 2^{\alpha+\frac{1}{2}}}{2^{\alpha+\frac{1}{2}}-1},
\end{aligned}$$

which indeed equals $D_\gamma F$.

Remark 3.4.21. *If one replaces a_p by $\widetilde{a}_p := 2^{-\frac{p}{2}}$ for $p \in \mathbb{N}_0$, then the theorem cannot be applied, $DF(t)$ is unbounded as $t \to 0$ and by analogous calculations as in the above example, one can verify that $D_\gamma F \neq \langle DF, \mathrm{d}\gamma\rangle$ even though both terms are finite.*

3.4.4 Rough and paracontrolled integrals

Compared to the whole history of integration, there has been a rather recent development beyond YS integration, namely rough integration. The motivation (or at least one out of several) was given by the desire to have integrals of the type $\int F(W)\mathrm{d}W$ for a Brownian Motion W, whose paths have only Hölder regularity $\alpha < \frac{1}{2}$. Its applicability to our setting seems to be limited by the discontinuity of the integrand H_{pm}. There exist results[12] stating under which conditions on $(\beta_{pm})_{(p,m)\in\Lambda_0}$ the series $\sum_{(p,m)\in\Lambda_0}\beta_{pm}H_{pm}(t)$ converges to a continuous function, but due to the complexity of both those conditions and the techniques of rough integration we decided to leave a thorough analysis into

[12]The survey paper [Gol73] presents an impressive number of results on Haar series of the structure $\sum_{(p,m)\in\Lambda_0} b_{pm}H_{pm}(t)$ ranging from convergence theorems for Haar series to the question of whether functions can be approximated by Haar series.

this direction for future research. Very likely, some kind of control of the $(\beta_{pm})_{(p,m)\in\Lambda_0}$ by the coefficients $(\gamma_{pm})_{(p,m)\in\Lambda_0}$ will be required for applicability of results of the type from [GIP15].

Remark 3.4.22. *If we were to define $\int_I DF(t)\mathrm{d}\gamma(t)$ analogously to* [GIP16] *as*

$$\int_I DF(t)\mathrm{d}\gamma(t) := \sum_{(p,m)\in\Lambda_0}\sum_{(q,n)\in\Lambda_0}\beta_{pm}\gamma_{qn}\int_I H_{pm}(t)\mathrm{d}G_{qn}(t) = \sum_{(p,m)\in\Lambda_0}\beta_{pm}\gamma_{pm},$$

then by construction we would have $D_\gamma F = \int_I DF(t,\cdot)\mathrm{d}\gamma(t)$. Already for $\gamma \in \mathcal{H}(I)$ we have seen that this property does not hold in general, hence such a definition would give different results than the ones we obtained above. In particular, a restriction to $\gamma \in \mathcal{H}(I)$ would yield, in the notation of [Øk97]*, $\mathbb{D}_{1,2} = \mathcal{D}_{1,2}$.*

3.5 Varying paths with continuous functions

In this Section it is a standing assumption that $\gamma \in C_0(I)$, i.e. the space of continuous functions on I vanishing in zero, with representation $\gamma(t) = \sum_{(p,m)\in\Lambda_0}\gamma_{pm}G_{pm}(t)$, $t \in I$.

3.5.1 Extensibility of Ciesielski's isomorphism

The proof of Theorem 3.4.8 cannot be directly extended (by letting $\alpha \to 0$ in the proof) to establish an isomorphism between ℓ^∞ and $C_0(I)$, because the bounds for the operator norm (3.4.2) become infinite as we let α get arbitrarily close to 0.

We recall that, by [Cie59, Theorem 3], any continuous function $x \in C(I)$ for $I = [0,1]$ has a Schauder representation

$$x(t) = x(0) + \sum_{(p,m)\in\Lambda_0}\left[\int_I H_{pm}(s)\mathrm{d}x(s)\right]G_{pm}(t),$$

but this representation is of little help unless the coefficients $(\int_I H_{pm}(s)\mathrm{d}x(s))_{(p,m)\in\Lambda_0}$ can be shown to belong to a nice sequence space. To illustrate that one can have a nice representation with bounded coefficients, we give the following example.

Example 3.5.1. *The function $\gamma\colon [0,1] \to \mathbb{R}$, given by $\gamma(0) := 0$ and $\gamma(t) := \frac{1}{\ln(t/2)}$ for $t \in (0,1]$, is (uniformly) continuous, but not Hölder continuous of any order $\alpha > 0$. We can verify this by verifying the (un-)boundedness of the sequence*

$$\xi^{(\alpha)}_{pm} := \int_0^1 \frac{2^{(p+1)\alpha}}{2\cdot\sqrt{2^p}}H_{pm}(t)\mathrm{d}\gamma(t), \quad (p,m) \in \Lambda_0.$$

To see that this sequence does not belong to ℓ^∞ for any positive α, we consider the subsequence with $m = 1$:

$$\begin{aligned}\left|\xi^{(\alpha)}_{p1}\right| &= \frac{2^{(p+1)\alpha}}{2}\left|2\gamma(t^1_{p1}) - \gamma(t^0_{p1}) - \gamma(t^2_{p1})\right| \\ &= \frac{2^{(p+1)\alpha}}{2}\left|\frac{2}{\ln(\frac{1}{2^{p+2}})} - \frac{1}{\ln(\frac{1}{2^{p+1}})}\right| \\ &= \frac{2^{(p+1)\alpha}\cdot p}{2\cdot\ln(2)\cdot(p+1)(p+2)},\end{aligned}$$

which is unbounded for $\alpha > 0$. *For* $\alpha = 0$, *however, the sequence* $\left(\xi_{pm}^{(\alpha)}\right)_{(p,m)\in\Lambda_0}$ *is bounded:*

$$\begin{aligned}
\left|\xi_{00}^{(0)}\right| &= \frac{1}{2}\left|\gamma(1) - \gamma(0)\right| = \frac{1}{2\ln 2},\\
\left|\xi_{pm}^{(0)}\right| &= \frac{1}{2}\left|2\gamma(t_{pm}^1) - \gamma(t_{pm}^0) - \gamma(t_{pm}^2)\right| \le \frac{1}{2}\left|4\cdot\gamma(1)\right| = \frac{2}{\ln 2}, \quad (p,m)\in\Lambda.
\end{aligned}$$

This example motivates an analysis of the maps corresponding to T_α and T_α^{-1} from Theorem 3.4.8 for $\alpha = 0$. To make this precise, let us first define the functions

$$\xi_{pm}^{(0)}(t) := \frac{1}{2\sqrt{2^p}}H_{pm}(t), \quad \phi_{pm}^{(0)}(t) := 2\sqrt{2^p}G_{pm}(t), \qquad \text{for } (p,m)\in\Lambda,\ t\in I$$

and

$$\xi_{00}^{(0)}(t) = \xi_{00}(t) \equiv 1 \quad \text{and} \quad \phi_{00}^{(0)}(t) = \phi_{00}(t) = t, \quad t\in I.$$

With these we can formulate a proposition.

Proposition 3.5.2. *The following mapping defines a bounded linear operator:*

$$\widetilde{T}\colon \begin{cases} (C_0(I), \|\cdot\|_\infty) & \longrightarrow (\ell^\infty, \|\cdot\|_{\ell^\infty}) \\ x & \longmapsto \left(\xi_{pm} := \int_0^1 \xi_{pm}^{(0)}(t)\mathrm{d}x(t)\right)_{(p,m)\in\Lambda_0} \end{cases}.$$

Proof. Linearity of $\widetilde{T}$ follows directly from the linearity of the integral with respect to the integrator. The upper bound is given by $\left\|\widetilde{T}\right\| \le 2$. To see this, calculate the absolute values:

- $|\xi_{00}| = |\int_0^1 1\mathrm{d}x(t)| \le |x(1)| + |x(0)| \le 2\|x\|_\infty$;
- $|\xi_{pm}| = |\int_0^1 \xi_{pm}^{(0)}(t)\mathrm{d}x(t)| \le \frac{1}{2}\left|2x(t_{pm}^1) - x(t_{pm}^0) - x(t_{pm}^2)\right| \le \frac{1}{2}\cdot 4\cdot\|x\|_\infty \le 2\|x\|_\infty$.

Hence $\left\|\widetilde{T}x\right\|_{\ell^\infty} \le 2\|x\|_\infty$, which proves the claim. □

Remark 3.5.3. *The proof can also be obtained from Lemma 2 in* [Cie59], *which states that if* ω_2 *denotes the second order modulus of continuity*[13] *of* $x\in C_0(I)$, *and if* $b_{pm} = \int_I H_{pm}(t)\mathrm{d}x(t)$ *for* $(p,m)\in\Lambda_0$, *then one has the estimate*

$$\sup_{1\le m\le 2^p}|b_{pm}| \le \sqrt{2^p}\omega_2\left(\frac{1}{2^p}\right),$$

from which we can infer that $\left(2^{-p/2}b_{pm}\right)_{(p,m)\in\Lambda_0} \in \ell^\infty$ *by dividing by* $\sqrt{2^p}$ *and taking the supremum over* $p\in\mathbb{N}_0$, *because for a continuous (and hence bounded) function* x *on* I, ω_2 *is bounded. In the above notation,* $\xi_{pm} = 2^{-(1+p/2)}b_{pm}$, *which proves the claim once the term* ξ_{00} *is added to the analysis.*

[13]For $x\in C([0,1])$, [Cie59] defines $\omega_2(\delta) := \sup\left\{ x(t_1) + x(t_2) - 2x(\frac{t_1+t_2}{2}) \;\middle|\; t_1,t_2\in[0,1], |t_1-t_2|\le\delta\right\}$ for $\delta\in[0,1]$.

The candidate for the inverse of $\widetilde{T}$,

$$T\colon \begin{cases} (\ell^\infty, \|\cdot\|_{\ell^\infty}) & \longrightarrow (C_0(I), \|\cdot\|_\infty) \\ \xi := (\xi_{pm})_{(p,m)\in\Lambda_0} & \longmapsto x := \sum_{(p,m)\in\Lambda_0} \xi_{pm}\phi^{(0)}_{pm}, \end{cases}$$

is not a bounded operator. It would be one, if its domain were ℓ^1 instead of ℓ^∞. To see this, let $x = T\xi$ for $\xi \in \ell^1$. Then we have for any $t \in I$

$$\begin{aligned} |x(t)| &= \left| \sum_{(p,m)\in\Lambda_0} \xi_{pm}\phi^{(0)}_{pm}(t) \right| \\ &\leq \sum_{(p,m)\in\Lambda_0} \left| 2\sqrt{2^p}\xi_{pm}G_{pm}(t^1_{pm}) \right| \\ &= \sum_{(p,m)\in\Lambda_0} \left| 2^{p+1}\xi_{pm}\left(t^1_{pm} - t^0_{pm}\right) \right| = \sum_{(p,m)\in\Lambda_0} |\xi_{pm}| = \|\xi\|_{\ell^1}\,. \end{aligned}$$

Thus, T would be bounded if its domain was ℓ^1, but that is strictly smaller than ℓ^∞. For instance, the sequence $\xi := (1, 1, \ldots) \in \ell^\infty$ yields $|(T\xi)(t)| = t + 2\sum_{(p,m)\in\Lambda} 2^{p/2}G_{pm}(t)$. Let $t_N := \sum_{n=0}^{N}(-1)^n(1/2)^n$. Then

$$\mathcal{G}_N := \sum_{n=0}^{\infty} \sqrt{2^n}G_{n1}(t_N) = \sum_{n=0}^{N} \sqrt{2^n}G_{n1}(t_N), \quad N \in \mathbb{N},$$

defines a strictly increasing and unbounded sequence (see Lemma B.3.2); therefore, $\sup_{N\in\mathbb{N}} |(T\xi)(t_N)| = \infty$, which proves the unboundedness of the operator T.

3.5.2 Directional derivative and integrated Malliavin derivative

The candidate for $D_\gamma F(\omega)$ is $\sum_{(p,m)\in\Lambda_0} \beta_{pm}(\omega) \int_I H_{pm}(t)\mathrm{d}\gamma(t)$. As before, we have to verify that the integral is well defined and that the sum converges if $F \in \mathcal{D}\backslash\mathcal{S}$. Furthermore we are again interested in the link to the Malliavin derivative.
Thus we ask

1. whether the Stieltjes integral $(\mathcal{R}) \int_I H_{pm}(t)\mathrm{d}\gamma(t)$ ($(p, m) \in \Lambda_0$) is well defined for a continuous, but not Hölder continuous γ;
2. whether the sum $\sum_{(p,m)\in\Lambda_0} \beta_{pm}(\omega) \int_I H_{pm}(t)\mathrm{d}\gamma(t)$ converges for fixed $\omega \in \Omega$ — in this case $D_\gamma F(\omega) = \sum_{(p,m)\in\Lambda_0} \beta_{pm}(\omega) \int_I H_{pm}(t)\mathrm{d}\gamma(t)$;
3. whether $DF(t, \omega) = \sum_{(p,m)\in\Lambda_0} \beta_{pm}(\omega)H_{pm}(t)$ is Stieltjes integrable w.r.t. γ on I;
4. under which conditions
$$\int_I DF(t, \omega)\mathrm{d}\gamma(t) = D_\gamma F(\omega).$$

The first two questions give us the directional derivative; the third question is a non-trivial extension of the first one asking whether DF is integrable against γ; the last question discusses again whether $D_\gamma F(\omega) = \langle DF(\cdot, \omega), \mathrm{d}\gamma\rangle$.

Denote by $BV(I)$ the functions of bounded (1-)variation on $I = [0,1]$.
With $H_{pm} \in BV(I)$ for $(p,m) \in \Lambda_0$ we can immediately answer the first question in the affirmative (combining Theorems B.1.5 and B.1.6). The other questions can be answered in one go as follows:

Theorem 3.5.4. *Fix $\omega \in \Omega$. Let $F \in \mathcal{D}$ be represented by f with $\nabla_{pm} f\left((\theta_k(\omega))_k\right) = \beta_{pm}(\omega)$. Assume that the sequence $(2^{p/2}\beta_{pm}(\omega))_{(p,m)\in\Lambda_0} \in \ell^1$. Let $\gamma \in C_0(I)$.*
Then $D_\gamma F(\omega) = \sum_{(p,m)\in\Lambda_0} \beta_{pm} \int_I H_{pm}(t)\mathrm{d}\gamma(t)$ and $DF \in \mathcal{R}(\gamma) \subset \mathcal{L}(\gamma)$, i.e., $D_\gamma F$ and $\int_I DF(t,\omega)\mathrm{d}\gamma(t)$ are well defined. Furthermore, one has the equality

$$\langle DF, \mathrm{d}\gamma\rangle = \int_I \lim_{N\to\infty} \sum_{(p,m)\in\Lambda_0^N} \beta_{pm} H_{pm}(t)\mathrm{d}\gamma(t) = \lim_{N\to\infty} \int_I \sum_{(p,m)\in\Lambda_0^N} \beta_{pm} H_{pm}(t)\mathrm{d}\gamma(t) = D_\gamma F.$$

Proof. For $(p,m) \in \Lambda_0$ let $\gamma_{pm} := \int_I H_{pm}(t)\mathrm{d}\gamma(t)$. From Proposition 3.5.2 we know that $(2^{-p/2}\gamma_{pm})_{(p,m)\in\Lambda_0} \in \ell^\infty$, hence, by Lemma B.2.7, $\sum_{(p,m)\in\Lambda_0} \beta_{pm}(\omega)\gamma_{pm}$ converges absolutely to $D_\gamma F(\omega)$. For proving the integrability of $DF(\cdot,\omega)$ w.r.t. γ in Stieltjes sense we follow our strategy from Section 3.4.3: By lower semi-continuity of the total variation (see Lemma B.2.11),

$$V_1(DF) \le \liminf_{N\to\infty} V_1(DF_N) \quad \text{for} \quad DF_N(\cdot,\omega) := \sum_{(p,m)\in\Lambda_0^N} \beta_{pm}(\omega)H_{pm}(\cdot).$$

Observe that each summand $\beta_{pm}H_{pm}$ adds at most $4\cdot 2^{p/2}\left|\beta_{pm}\right|$ to the 1-variation, hence

$$V_1(DF) \le 4 \sum_{(p,m)\in\Lambda_0} 2^{p/2}\left|\beta_{pm}\right|.$$

If the above expression is finite, $DF(\cdot,\omega)$ has finite variation. A sufficient condition is that the sequence $\left(2^{p/2}\beta_{pm}(\omega)\right)_{(p,m)\in\Lambda_0} \in \ell^1$, which is precisely what we assumed. With this we know that $\langle DF, \mathrm{d}\gamma\rangle = \int_I DF(t,\omega)\mathrm{d}\gamma(t)$ is also well defined (as RS integral and hence also as LS integral). For exchanging integral and limit we want to apply the dominated convergence theorem for LS integrals (see Theorem B.1.12). In this situation, $\sum_{(p,m)\in\Lambda_0^N} \beta_{pm}(\omega)H_{pm}(\cdot)$ is dominated by $|DF(\cdot,\omega)|$, which is, as we already saw, integrable w.r.t. γ. By Corollary B.2.9, the series converges absolutely and uniformly to $DF(\cdot,\omega)$ on I as $N\to\infty$. Thus dominated convergence gives the desired result. □

3.5.3 The path ahead

Throughout we worked with $F\colon \Omega \to \mathbb{R}$ where $\Omega = C(I)$ and the object of interest was $D_\gamma F(\omega) := \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left[F(\omega+\varepsilon\gamma) - F(\omega)\right]$. As long as γ belongs to some subset of Ω, that definition makes perfect sense. While our paths $\omega \in \Omega$ are still continuous, we now want to perturb these paths in a discontinuous manner. Provided that the integrals $\theta_{pm}(\gamma) = \int_I H_{pm}(t)\mathrm{d}\gamma(t)$ have a meaning for $\gamma \notin C(I)$, there is nothing stopping us from allowing such $\widetilde{\omega}$ as argument of F. Our ultimate goal will be to give a meaning to $D_\gamma F(\omega)$ in such a way that will allow us to recover the vertical derivative from [Dup09] and [CF13].

3.6 Varying paths with discontinuous functions

Before we can look at the directional derivative, we first have to extend the domain of our functional F beyond the continuous functions on I. To see what would be a suitable domain, we must first state the type of integration we are considering. Only then will we ask the same questions as before on existence of a well defined directional derivative, integrability of the Malliavin derivative against the chosen direction and when the directional derivative $D_\gamma F$ has the representation $\langle DF, \mathrm{d}\gamma\rangle$. It is clear that RS integration is not suitable here, because it does not allow integrand and integrator to have common points of discontinuity. Instead, we will consider the LS integral and the generalized RS integral.

3.6.1 Generalized Riemann-Stieltjes integral

On page 194 in [McL80] one can find the following statement (adapted to our setting):

Lemma 3.6.1. *Suppose f is a function of bounded variation on I. Then $(\mathcal{G})\int_I f(t)\mathrm{d}\gamma(t)$ exists when γ is a regulated function.*

The integration by parts formula holds, allowing to also define the integral for regulated integrand and integrator with bounded variation; see Theorem B.1.7.

Haar functions are clearly functions of bounded variation, hence for any regulated function γ on I and for any $(p, m) \in \Lambda_0$, the integral $(\mathcal{G})\int_I H_{pm}\mathrm{d}\gamma$ exists. For the directional derivative $D_\gamma F$, which, if it exists equals $\sum_{(p,m)\in\Lambda_0} \beta_{pm} \int_I H_{pm}(t)\mathrm{d}\gamma(t)$, we still need to ensure the convergence of that series.

Furthermore, $DF(\cdot, \omega)$, being the uniform limit (under conditions given in Corollary B.2.9) of the step functions $\sum_{(p,m)\in\Lambda_0^N} \beta_{pm}(\omega)H_{pm}(t)$ as $N \to \infty$, is certainly a regulated function. Thus it is integrable against $\gamma \in BV(I)$. If $DF(\cdot, \omega)$ has bounded variation (for which we have derived criteria in previous sections), then its integrability is given for any regulated integrator γ.

Thus we can consider directional derivatives $D_\gamma F(\omega)$ for a fixed $\omega \in \Omega$ and for γ being a regulated function and the further analysis of the link to the Malliavin derivatives will require either $DF(\cdot, \omega)$ or γ to have bounded variation. In the following subsection we present the different cases mentioned here in detail.

Varying paths with functions of bounded variation

If we consider functions on the extended domain $BV(I)$ belonging to the space

$$\mathcal{D}^{BV} := \left\{ F\colon BV(I) \to \mathbb{R} \,\middle|\, F(\omega) = f\left((\theta_k(\omega))_{k\in\mathbb{N}}\right),\ f \in C^\infty(\mathbb{R}^\mathbb{N}),\ (\beta_k(\omega))_{k\in\mathbb{N}} \in \ell^2 \right\},$$

then we can define the directional derivative for fixed $\omega \in \Omega$ and $\gamma \in BV(I)$,

$$D_\gamma F(\omega) := \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left[F(\omega + \varepsilon\gamma) - F(\omega)\right],$$

provided the limit exists. From previous calculations we know that if this limit exists, it must equal $\sum_{(p,m)\in\Lambda_0} \beta_{pm}(\omega)\gamma_{pm}$ where we recall the notation for $(p, m) \in \Lambda_0$ and $\omega \in \Omega$:

$$\beta_{pm}(\omega) := \nabla_{pm} f\left((\theta_{qn}(\omega))_{(q,n)\in\Lambda_0}\right), \qquad \gamma_{pm} := \int_I H_{pm}(t)\mathrm{d}\gamma(t).$$

The integral defining γ_{pm} exists in generalized RS sense, because Haar functions are step functions and thus regulated.

Theorem 3.6.2. *Fix $\omega \in \Omega$ and let $F \in \mathcal{D}^{BV}$ and $\gamma \in BV(I)$. If $(2^{p/2}\beta_{pm}(\omega))_{(p,m)\in\Lambda_0} \in \ell^1$, then $DF(t,\omega) = \sum_{(p,m)\in\Lambda_0} \beta_{pm}(\omega)H_{pm}(t)$ is RS integrable in the generalized sense w.r.t. γ. Furthermore, $D_\gamma F(\omega)$ exists and equals $(\mathcal{G})\int_I DF(t,\omega)\mathrm{d}\gamma(t)$.*

Proof. From Corollary B.2.9 we infer that $DF(\cdot,\omega)$ is regulated and therefore integrable against an integrator with bounded variation. Furthermore, $DF(\cdot,\omega)$ is the uniform limit of integrable step functions; therefore, by uniform convergence (Theorem B.1.9),

$$(\mathcal{G})\int_I DF(t,\omega)\mathrm{d}\gamma = \lim_{N\to\infty} \sum_{(p,m)\in\Lambda_0^N} \beta_{pm}(\omega)\gamma_{pm} = D_\gamma F(\omega),$$

where the second equality holds by definition of the directional derivative. □

Remark 3.6.3 (complementing the proof)**.** *One can check directly that the series defining $D_\gamma F(\omega)$ converges: Any function of bounded variation on I is trivially bounded, hence $|\gamma_{pm}| \le \sqrt{2^p}\,\|\gamma\|_\infty$. Thus $(\beta_{pm}(\omega)\gamma_{pm})_{(p,m)\in\Lambda_0} \in \ell^1$ by Lemma B.2.7.*

Let us remark that if the integrator γ is a discontinuous function and we consider $\int_I H_{pm}(t)\mathrm{d}\gamma(t)$ for some $(p,m)\in\Lambda$, then the precise (pointwise) definition of H_{pm} matters. To illustrate this, we calculate that for a nondecreasing $\gamma\colon I\to\mathbb{R}$ and $m \notin \{1,2^p\}$

$$\begin{aligned}\int_I H_{pm}(t)\mathrm{d}\gamma(t) = \sqrt{2^p}\Big[&\frac{1}{2}\left(\gamma(t^0_{pm+}) - \gamma(t^0_{pm-})\right) - \frac{1}{2}\left(\gamma(t^2_{pm+}) - \gamma(t^2_{pm-})\right)\\ &+ \gamma(t^1_{pm-}) - \gamma(t^0_{pm+}) - \gamma(t^2_{pm-}) + \gamma(t^1_{pm+})\Big],\end{aligned} \tag{3.6.1}$$

where $\gamma(t_-)$ and $\gamma(t_+)$ denote left and right limit of γ in $t\in I$. The expression in (3.6.1) is sensitive to the exact value of the Haar function at t^0_{pm} and t^2_{pm} unless γ is continuous in those points.[14]

Example 3.6.4. *Let $p\in\mathbb{N}$ and $m\in\{2,\dots,2^p-1\}$. For such values consider the integral $\int_I H_{pm}(t)\mathrm{d}\gamma(t)$ when $\gamma\colon I\to\mathbb{R}$ is the shifted Heaviside function $\gamma^H_x(t) := \mathbb{1}_{[x,1]}(t)$ for $x\in I$. By applying* (3.6.1) *one can verify that indeed for any $t\in I$, $\int_I H_{pm}(t)\mathrm{d}\gamma^H_x(t) = H_{pm}(x)$.*

Varying paths with regulated functions

We extend the domain from Ω even further to the regulated functions $\mathfrak{R}(I)$ and consider the space

$$\mathcal{D}^R := \left\{ F\colon \mathfrak{R}(I)\to\mathbb{R} \,\middle|\, F(\omega) = f\left((\theta_k(\omega))_{k\in\mathbb{N}}\right),\ f\in C^\infty(\mathbb{R}^{\mathbb{N}}),\ (\beta_k(\omega))_{k\in\mathbb{N}}\in\ell^2\right\}.$$

Analogously to the previous section we observe that for such functionals, the directional derivative $D_\gamma F(\omega)$ for fixed $\omega\in\Omega$ and $\gamma\in\mathfrak{R}(I)$ exists and equals $\sum_{(p,m)\in\Lambda_0}\beta_{pm}(\omega)\gamma_{pm}$, provided that this sum converges. For the integrability of $DF(\cdot,\omega)$ against $\gamma\in\mathfrak{R}(I)$, we require $DF(\cdot,\omega)\in BV(I)$.

[14]The restricting choice of values for p and m is such that the interval $[t^0_{pm}, t^2_{pm}]$ lies strictly included in I. This prevents a tedious distinction of different cases depending on whether the endpoints of the interval coincide with either 0 or 1. Such a distinction would only complicate the formulas without adding further insights, hence we omit it here.

Theorem 3.6.5. *Fix $\omega \in \Omega$ and let $F \in \mathcal{D}^R$ and $\gamma \in \mathfrak{R}(I)$. Assume that*

$$(2^{p/2}\beta_{pm}(\omega))_{(p,m)\in\Lambda_0} \in \ell^1.$$

Then $D_\gamma F(\omega)$ exists. Furthermore, $DF(\cdot,\omega) \in BV(I)$, hence $\int_I DF(t,\omega)\mathrm{d}\gamma(t)$ exists.

Proof. γ is regulated on a compact interval, hence it is bounded. This fact, combined with the condition $(2^{p/2}\beta_{pm}(\omega))_{(p,m)\in\Lambda_0} \in \ell^1$, allows us to argue as in Remark 3.6.3 that the series defining $D_\gamma F(\omega)$ converges (absolutely). As we have seen in Theorem 3.5.4, the condition also implies that $DF(\cdot,\omega) \in BV(I)$, hence the generalized RS integral $\int_I DF(t,\omega)\mathrm{d}\gamma(t)$ exists (and is finite). □

The theorem for uniform convergence of Stieltjes integrals (Theorem B.1.9) cannot be applied, because the integrator γ is not of bounded variation. This property is essential for the proof of the theorem. Therefore, equality of $\int_I DF(t,\omega)\mathrm{d}\gamma(t)$ and $D_\gamma F(\omega)$ does not follow in this case.

3.6.2 Lebesgue-Stieltjes integral

The LS integral $(\mathcal{L})\int_I f(t)\mathrm{d}g(t)$ can be introduced either as the abstract Lebesgue integral $(\mathcal{L})\int_I f(t)\mathrm{d}\mu_g(t)$ when μ_g is the LS measure associated to g, or as an extension of the RS integral. We quickly go over the latter approach before going into details with the former approach. The LS integral is an interim step towards the generalized RS integral, which we considered already. Hence, the LS integral will not yield any results going beyond what we have already seen. The reason is that we integrate w.r.t. functions of bounded variation, but not w.r.t. the larger class of regulated functions. Therefore, we will keep the integrability results as short as possible and devote more time to the discussion of measures.

Varying paths with functions of bounded variation

The LS integral on I, $(\mathcal{L})\int_I f(t)\mathrm{d}g(t)$, is defined if f is measurable and bounded and g is of bounded variation and right-continuous.[15]

Thus we can define $\gamma_{pm} := (\mathcal{L})\int_I H_{pm}(t)\mathrm{d}\gamma(t)$ for γ being right-continuous and of bounded variation. This permits us to write down the corresponding directional derivative $D_\gamma F(\omega)$ for $F \in \mathcal{S}$ or $\mathcal{D}$ and to analyze the convergence of the series defining $D_\gamma F(\omega)$. For the integrability of $DF(\cdot,\omega)$ w.r.t. such an integrator γ we have to ensure that $DF(\cdot,\omega)$ is Borel-measurable and bounded. For $F \in \mathcal{S}$ both properties are satisfied. For $F \in \mathcal{D}$ we have to establish the appropriate conditions.

Theorem 3.6.6. *Fix $\omega \in \Omega$ and let $F \in \mathcal{D}^{BV}$ with $(2^{p/2}\beta_{pm}(\omega))_{(p,m)\in\Lambda_0} \in \ell^1$. Let $\gamma \in BV(I)$ be right-continuous with $\gamma_{pm} := (\mathcal{L})\int_I H_{pm}(t)\mathrm{d}\gamma(t)$ for $(p,m) \in \Lambda_0$.*
Then $D_\gamma F(\omega) = \sum_{(p,m)\in\Lambda_0}\beta_{pm}(\omega)\gamma_{pm}$ is an absolutely convergent series, $DF(\cdot,\omega)$ is integrable w.r.t. γ and $D_\gamma F(\omega) = (\mathcal{L})\int_I DF(t,\omega)\mathrm{d}\gamma(t)$.

[15] See for instance [KK68, Section 4.6-17] for the definition of the LS-integral.

Proof. We have seen in the proof of Theorem 3.5.4 that under the above conditions, $DF(\cdot,\omega) \in BV(I)$, hence it is measurable and bounded and therefore integrable. It is also the uniform limit of step functions, hence we have once more

$$(\mathcal{L}) \int_I DF(t,\omega)\mathrm{d}\gamma = \lim_{N\to\infty} \sum_{(p,m)\in\Lambda_0^N} \beta_{pm}(\omega)\gamma_{pm} = D_\gamma F(\omega).$$

The absolute convergence of the series can be checked as follows: $\gamma \in BV(I)$ is in particular bounded, hence for $(p,m) \in \Lambda_0$ we have

$$|\beta_{pm}(\omega)\gamma_{pm}| \leq \left|2^{p/2}\beta_{pm}(\omega)\right| \cdot \sup_{(p,m)\in\Lambda_0} |\gamma_{pm}|,$$

which defines an ℓ^1-sequence by Lemma B.2.7. □

3.7 Varying paths with measures

3.7.1 Lebesgue-Stieltjes signed measures

To every nondecreasing right-continuous[16] function γ on I one can associate the LS measure μ_γ such that γ is then the distribution function of μ_γ.[17] Using the (unique[18]) Jordan decomposition (Theorem B.2.15) of functions of bounded variation one can extend this construction. Thus we can define a *LS signed measure* associated to a function of bounded variation.

Definition 3.7.1 (Lebesgue-Stieltjes signed measure, cf. [CK04, Definition 7.28]). *Let $\gamma \in BV(I)$ with Jordan decomposition $\gamma = \gamma_1 - \gamma_2$, where γ_1 and γ_2 are both nondecreasing functions on I. Let μ_1 and μ_2 denote the (finite) LS measures associated to γ_1 and γ_2, respectively. Then $\mu_\gamma := \mu_1 - \mu_2$ defines a σ-additive set function on $\mathfrak{B}(I)$. We call μ_γ the LS signed measure associated to γ.*

On the other hand, if μ is a LS measure, then we can associate a function g_μ to this measure. With the terminology of probability theory (though μ need not be a probability measure), γ_μ is the (cumulative) distribution function associated to μ, i.e. $\gamma_\mu(t) := \mu((-\infty,t])$. Then we can interpret $D_\mu F(\omega)$ as the directional derivative that we obtain from perturbing ω in direction γ_μ. This approach can be extended to the LS signed measure corresponding to functions of bounded variation. This correspondence between functions and (signed) measures motivates the introduction of a directional derivative when the direction is a finite measure or signed measure on I.

Definition 3.7.2. *Let $F \in \mathcal{D}^{BV}$ and μ a finite LS measure on I with associated distribution function γ_μ. Then we define the directional derivative of $F(\omega)$ in direction μ as $D_\mu F(\omega) := D_{\gamma_\mu} F(\omega)$, provided the latter exists.*

The existence of $D_{\gamma_\mu} F$ and its link to the Malliavin derivative were discussed in Section 3.6.2 and can be transferred to this setting. In the next section we look at the integral defined in the abstract Lebesgue sense.

[16] If γ is not right-continuous, then one defines the LS measure by associating it to the right-continuous version of γ.

[17] See e.g. [CK04, Section 7.3.1] for the construction of the LS measure.

[18] See [CK04, Section 7.3.3] on the construction of the (minimal) Jordan decomposition of a function of bounded variation.

3.7.2 The integral defined with respect to a measure

There is no need to always go back to the Stieljes-type integrals. Alternatively one can work with abstract Lebesgue integrals as they are defined for instance in [Nie97, Section 13.2]. Haar functions are step functions on dyadic intervals, hence they are measurable. We can use the dominated convergence theorem (e.g. [Nie97, Theorem 13.15]) to obtain integrability of DF against $\mu \in \mathfrak{M}(I)$. A link between the LS integral and the abstract Lebesgue integral can be found in [Nie97, Section 13.4].

A candidate for the directional derivative of $F \in \mathcal{D}$ with (infinite) representation $F(\omega) = f\left((\theta_{pm}(\omega))_{(p,m)\in\Lambda_0}\right)$ in direction of a measure μ, provided the series converges, is

$$D_\mu F(\omega) := \sum_{(p,m)\in\Lambda_0} \beta_{pm}(\omega) \int_I H_{pm}(t)\mathrm{d}\mu(t), \tag{3.7.1}$$

where $\beta_{pm}(\omega) = \nabla_{pm} f((\theta_{qn}(\omega))_{(q,n)\in\Lambda_0})$. We calculate the integral for $(p,m) \in \Lambda$:

$$\begin{aligned} \mu_{pm} :=& \int_I H_{pm}(t)\mathrm{d}\mu(t) \\ =& \sqrt{2^p}\left[\mu((t^0_{pm}, t^1_{pm})) - \mu((t^1_{pm}, t^2_{pm}))\right] \\ &+ \sqrt{2^p}\left[\frac{1}{2}\mu(\{t^0_{pm}\}) - \frac{1}{2}\mu(\{t^2_{pm}\}) + \frac{1}{2}\mu(\{0\})\delta_{m1} - \frac{1}{2}\mu(\{1\})\delta_{m2^p}\right] \end{aligned} \tag{3.7.2}$$

and put $\mu_{00} := \int_I H_{00}(t)\mathrm{d}\mu(t) = \mu(I)$.

Remark 3.7.3. *If we had chosen to work with Definition 3.4.4 of Haar functions, we would instead have*

$$\widetilde{\mu}_{pm} := \int_I \widetilde{H}_{pm}(t)\mathrm{d}\mu(t) = \sqrt{2^p}\left[\mu([t^0_{pm}, t^1_{pm})) - \mu([t^1_{pm}, t^2_{pm}))\right] \quad \textit{for } (p,m) \in \Lambda.$$

In the special case of the Lebesgue measure $\mu = \lambda|_I$, one has $\mu_{00} = \lambda(I) = 1$ and $\mu_{pm} = 0$ for all $(p,m) \in \Lambda$. In the special case of the Dirac measure $\mu = \delta_t$ one has $\mu_{pm} = H_{pm}(t)$ for $(p,m) \in \Lambda_0$.

So far we always let $D_\mu F(\omega) := \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left[F(\omega + \varepsilon\mu) - F(\omega)\right]$. In the next Section we say how we can give a meaning to this if $F \in \mathcal{D}$, $\omega \in \Omega = C(I;\mathbb{R})$ and $\mu \in \mathfrak{M}(I)$.

3.7.3 Variation of a path with a measure

Recall that for $\omega \in \Omega$ and a Haar function H_{pm} ($(p,m) \in \Lambda_0$) we defined

$$\theta_{pm}(\omega) := \int_I H_{pm}(t)\mathrm{d}\omega(t).$$

For a measure $\mu \in \mathfrak{M}(I)$ we now define analogously[19]

$$\vartheta_{pm}(\mu) := \int_I H_{pm}(t)\mathrm{d}\mu(t) = \mu_{pm}.$$

[19]We use the notation $\int_I f(t)\mathrm{d}\mu(t)$ instead of $\int_I f(t)\mu(\mathrm{d}t)$ for the integral w.r.t. a measure μ.

This is well defined because Haar functions are simple functions and trivially belong to $\mathcal{M}(I)$. We had already observed that for functions $\omega, \nu \in \Omega$ and indices $(p,m) \in \Lambda_0$ one has $\theta_{pm}(\omega + \varepsilon\nu) = \theta_{pm}(\omega) + \varepsilon\theta_{pm}(\nu)$, i.e., $\theta_{pm} \colon \Omega \to \mathbb{R}$ is a linear functional. The same is true for $\vartheta_{pm} \colon \mathfrak{M}(I) \to \mathbb{R}$ introduced above. For $\mu, \widehat{\mu} \in \mathfrak{M}(I)$, the integrals $\int_I H_{pm} \mathrm{d}\mu$ and $\int_I H_{pm} \mathrm{d}\widehat{\mu}$ exist. Furthermore, $\mu + \varepsilon\widehat{\mu} \in \mathfrak{M}(I)$ and H_{pm} is integrable w.r.t. $\mu + \varepsilon\widehat{\mu}$ and we have $\vartheta_{pm}(\mu + \varepsilon\widehat{\mu}) = \vartheta_{pm}(\mu) + \varepsilon\vartheta_{pm}(\widehat{\mu})$. With this notation, the directional derivative in direction $\gamma \in C_0(I)$ could be written as

$$D_\gamma F(\omega) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left[f\left((\theta_{pm}(\omega) + \varepsilon\theta_{pm}(\gamma))_{(p,m)\in\Lambda_0} \right) - F(\omega) \right].$$

Analogously we define the directional derivative of $F(\cdot) = f((\theta_{pm}(\cdot))_{pm}) \in \mathcal{D}$ in direction $\mu \in \mathfrak{M}(I)$ as

$$D_\mu F(\omega) := \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left[f\left((\theta_{pm}(\omega) + \varepsilon\vartheta_{pm}(\mu))_{(p,m)\in\Lambda_0} \right) - F(\omega) \right],$$

provided the limit exists. Let us first make clear that this definition is indeed reasonable by considering an LS measure μ_g associated to a function $g \in BV(I)$ such that $\int_I H_{pm}(t) \mathrm{d}g(t) = \int_I H_{pm}(t) \mathrm{d}\mu_g(t)$ for any Haar function. Then clearly $D_g F(\omega) = D_{\mu_g} F(\omega)$ for all $\omega \in \Omega$, hence this definition is equivalent to Definition 3.7.2. If the limit in the definition exists, then we must have (in analogy with (3.3.1))

$$D_\mu F(\omega) = \sum_{(p,m)\in\Lambda_0} \beta_{pm}(\omega)\vartheta_{pm}(\mu).$$

If $F \in \mathcal{S}$, this sum is finite, hence nothing more needs to be done. If $F \in \mathcal{D}$, the convergence of this series is an issue to be addressed.

Conditions for convergence of the series and existence of $D_\mu F$

As before, for $F \in \mathcal{D}$ write $\beta_{pm}(\omega) := \nabla_{pm} f\left((\theta_{qn}(\omega))_{(q,n)\in\Lambda_0} \right)$ and $\mu_{pm} := \vartheta_{pm}(\mu)$ (see (3.7.2)). For fixed $\omega \in \Omega$ we want to find sufficient conditions for the (pointwise) convergence of $\sum_{(p,m)\in\Lambda_0} \beta_{pm}(\omega)\mu_{pm}$. From the definition of $\mathcal{D}$ we have that $(\beta_{pm})_{(p,m)\in\Lambda_0} \in \ell^2$. We can either impose further restrictions on F or, as previously done, require $(\mu_{pm})_{(p,m)\in\Lambda_0} \in \ell^2$.

- If μ has Lebesgue density $f_\mu \in C(I)$, then we can apply the previously developed theory to $\mu_{pm} = \int_I H_{pm}(t) \mathrm{d}\mu(t) = \int_I H_{pm}(t) f_\mu(t) \mathrm{d}t$.
- If we let $\mu := \delta_t$ the Dirac measure, then we already saw that $\mu_{pm} = H_{pm}(t)$, which defines for $(p,m) \in \Lambda_0$ an unbounded sequence; in particular $(\mu_{pm})_{(p,m)\in\Lambda_0} \notin \ell^2$. If we multiply this sequence by $2^{-p/2}$, then the sequence becomes bounded with $\left\| (2^{-p/2}\mu_{pm})_{(p,m)\in\Lambda_0} \right\|_{\ell^\infty} = 1$. Hence if $(2^{p/2}\beta_{pm})_{(p,m)\in\Lambda_0} \in \ell^1$, then $(\beta_{pm}\mu_{pm})_{(p,m)\in\Lambda_0} \in \ell^1$, i.e., $D_\mu F(\omega) := \sum_{(p,m)\in\Lambda_0} \beta_{pm}(\omega)\mu_{pm}$ is well defined.

Inspired by the above thoughts, we formulate our result:

Theorem 3.7.4. *Let $\mu \in \mathfrak{M}(I)$ denote a bounded measure with $\mu(I) =: M < \infty$. If $(2^{p/2}\beta_{pm})_{(p,m)\in\Lambda_0} \in \ell^1$, then $D_\mu F(\omega)$ is given by the value of the absolutely converging series $\sum_{(p,m)\in\Lambda_0} \beta_{pm}(\omega)\mu_{pm}$.*

Proof. From the boundedness we infer that

$$|\mu_{pm}| \leq \sqrt{2^p}\left[\mu([t^0_{pm}, t^1_{pm})) + \mu((t^1_{pm}, t^2_{pm}])\right] \leq \sqrt{2^p} \cdot M,$$

hence $\left\|(2^{-p/2}\mu_{pm})_{(p,m)\in\Lambda_0}\right\|_{\ell^\infty} \leq M$. Application of Lemma B.2.7 to the bounded sequence $(2^{-p/2}\mu_{pm})_{(p,m)\in\Lambda_0}$ and the ℓ^1-sequence $(2^{p/2}\beta_{pm})_{(p,m)\in\Lambda_0}$ yields the desired result. □

Integrability of DF against $\mu \in \mathfrak{M}(I)$

Fix $\omega \in \Omega$. If $F \in \mathcal{S}$, then $DF(t,\omega) = \sum_{(p,m)\in\Lambda_0^N} \beta_{pm}(\omega)H_{pm}(t)$ for some $N \in \mathbb{N}$. As $H_{pm} \in \mathcal{M}(I)$, so is $DF(\cdot,\omega)$, and the integral $\int_I DF(t,\omega)\mathrm{d}\mu(t)$ is well defined. For $F \in \mathcal{D}$, we observe that again $DF(t,\omega) = \sum_{(p,m)\in\Lambda_0} \beta_{pm}(\omega)H_{pm}(t)$ belongs to $\mathcal{M}(I)$[20]. For rightfully writing $\langle DF, \mathrm{d}\mu\rangle = \int_I DF(t,\omega)\mathrm{d}\mu(t)$ we need to verify the integrability of $DF(\cdot,\omega)$ against the measure μ.

Theorem 3.7.5. *Fix $\omega \in \Omega$. Let $\mu \in \mathfrak{M}(I)$ and assume $(2^{p/2}\beta_{pm}(\omega))_{(p,m)\in\Lambda_0} \in \ell^1$. Then $DF(\cdot,\omega)$ is regulated and integrable against μ, i.e., $\int_I DF(t,\omega)\mathrm{d}\mu(t)$ exists and is finite.*

Proof. We have already seen that the condition on $(\beta_{pm}(\omega))_{(p,m)\in\Lambda_0}$ implies that $DF(\cdot,\omega)$ is the uniform limit of step functions, hence regulated; therefore, $DF(\cdot,\omega)$ is integrable w.r.t. μ. Both $DF(\cdot,\omega)$ and μ are finite, hence the integral is also finite. □

Is $D_\mu F = \langle DF, \mathrm{d}\mu\rangle$?

As the convergence to DF is uniform in the previous theorems, we can immediately infer from the uniform convergence theorem for the abstract Lebesgue integral (w.r.t. finite (signed) measures[21]) the following result, which we therefore state without repeating the arguments of the proof:

Corollary 3.7.6. *Fix $\omega \in \Omega$. Let $\mu \in \mathfrak{M}(I)$ and assume $(2^{p/2}\beta_{pm}(\omega))_{(p,m)\in\Lambda_0} \in \ell^1$. Then $D_\mu F(\omega)$ and $\int_I DF(t,\omega)\mathrm{d}\mu(t)$ exist and are equal.*

Remark 3.7.7. *One could also argue via dominated convergence as follows: The finite sum $\sum_{(p,m)\in\Lambda_0^N} \beta_{pm}(\omega)H_{pm}(t)$ is integrable for any finite measure $\mu \in \mathfrak{M}(I)$. It is dominated by $|DF(\cdot,\omega)|$. Assuming that $DF(\cdot,\omega)$ is integrable, this also implies that $|DF(\cdot,\omega)|$ is integrable (per definition of the abstract Lebesgue integral), hence it is a dominating integrable function; the dominated convergence theorem for the abstract Lebesgue integral (cf.* [Nie97, Theorem 13.15] *or Theorem B.1.11 in the Appendix) gives the desired result.*

3.7.4 Link to the vertical derivative in Cont & Fournié

Summary of Results by Cont & Fournié

Let us briefly review the definitions from [CF13]:

[20] See for instance Section 10 in [Nie97], in particular Propositions 10.7 and 10.9 and Corollary 10.11, for the measurability of a (countably infinite) sum of measurable functions.

[21] The uniform convergence theorem for finite measures (Theorem B.1.10) can be extended to finite signed measure by applying the Jordan decomposition theorem (Theorem B.2.15).

Let $X\colon [0,T] \to \mathbb{R}^d$ be a continuous semimartingale on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ with paths in $C_0([0,T], \mathbb{R}^d)$. Let $[X](t) = \int_0^t A(u)\mathrm{d}u$ and let $Y(t) = F_t(X_t, A_t)$ depend on the entire paths of X and A, where $X_t = (X(u), u \in [0,t])$.

Introduce the notations for horizontal extension and vertical perturbation of a process $x_t = (x(u), 0 \le u \le t)$, which we illustrate in Figure 3.2:

$$
\begin{aligned}
x_{t,h}(u) &= x(u), \quad u \in [0,t]; \qquad & x_{t,h}(u) &= x(t), \quad u \in (t, t+h]; \\
x_t^h(u) &= x(u), \quad u \in [0,t); \qquad & x_t^h(t) &= x(t) + h.
\end{aligned}
$$

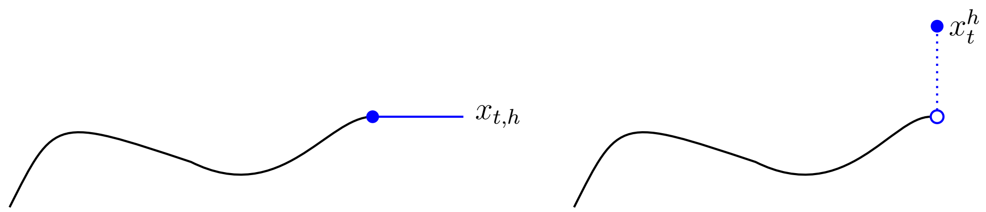


Figure 3.2: Horizontal extension $x_{t,h}$ and vertical perturbation x_t^h of a path x

With this the *horizontal derivative* is defined as

$$\mathcal{D}_t F(x,v) := \lim_{h \to 0^+} \frac{1}{h} \left(F_{t+h}(x_{t,h}, v_{t,h}) - F_t(x_t, h_t) \right)$$

and the *vertical derivative* is defined as

$$\nabla_x F(x,v) := (\partial_i F_t(x,v), i = 1, \ldots, d) \text{ where } \partial_i F_t(x,v) := \lim_{h \to 0} \frac{1}{h} \left(F_t(x_t^{he^i}, v) - F_t(x,v) \right),$$

if the limits exist and where $(e_i, i = 1, \ldots, d)$ is the canonical basis in $\mathbb{R}^d$.

Remark 3.7.8. *Horizontal (w.r.t. time) and vertical (w.r.t. the state space) differentiation were already introduced by Dupire in* [Dup09] *and also appeared in* [CF10].

The following results are obtained in [CF13]: There is a change of variable formula (using both $\mathcal{D}_t F$ and $\nabla_x F$). The vertical derivative of $Y(t) = F_t(X_t, A_t)$, $\nabla_X Y(t) := \nabla_x F_t(X_t, A_t)$, is independent of the representation F, it has a strong link to the Malliavin derivative, and ∇_X is the inverse of the Itô integral, i.e., $\nabla_X(\int \phi \cdot \mathrm{d}X) = \phi$.

Examples of both vertical and horizontal derivatives can be found in [CF13, Section 3.1, pp. 117-118].

Recently, horizontal and vertical derivatives have been nicely explained in the second part of [BCC16]. See in particular Table 7.1 therein, where R. Cont compares the property of the Malliavin derivative and the vertical derivative.

Link to our directional derivative in direction of a measure

For $t \in I$ consider the (stopped) trajectory $\omega_t = \{\omega(s) \mid s \in [0,t]\}$. Then we have $\theta_{pm}(\omega_t) = \int_0^t H_{pm}(s)\mathrm{d}\omega(s)$. Fix again $F \in \mathcal{D}$ with $F(\omega) = f\left((\theta_{pm}(\omega))_{(p,m)\in\Lambda_0}\right)$. For this we introduce the notation $\beta_{pm}(\omega_t) = \nabla_{pm} f\left((\theta_{qn}(\omega_t))_{(q,n)\in\Lambda_0}\right)$. The directional derivative of F in direction δ_t, evaluated in ω_t, can be calculated as

$$\begin{aligned} D_{\delta_t} F(\omega_t) &= \sum_{(p,m)\in\Lambda_0} \beta_{pm}(\omega_t)\vartheta_{pm}(\delta_t) = \sum_{(p,m)\in\Lambda_0} \beta_{pm}(\omega_t)\int_I H_{pm}(s)\delta_t(\mathrm{d}s) \\ &= \sum_{(p,m)\in\Lambda_0} \beta_{pm}(\omega_t) H_{pm}(t). \end{aligned}$$

On the other hand, if we drop the dependence on the quadratic variation process and let $d = 1$ in [CF13], then

$$\nabla_\omega F(\omega) = \partial_1 F_t(\omega) = D_{\delta_t} F(\omega_t),$$

where F_t shall denote the restriction of F to the stopped path ω_t. Thus our directional derivative in direction of δ_t for F_t yields exactly the vertical derivative. Instead of the Dirac measure we could also have chosen the Heaviside function as direction, which appeared in Example 3.6.4. This is no surprise as we already stated that the correspondence between finite measures and increasing (right-continuous) functions is one-to-one.

Remark 3.7.9. *One can see clearly that $D_{\delta_t} F(\omega_t)$ requires a pointwise exact definition of the Haar function. Definitions 3.4.3 and 3.4.4 yield different results here, which is why we drew attention to this issue in the first place. If one chose to work with differentiation in a weak (distributional) sense, then this difference would vanish.*

3.8 Conclusion

We have seen that for $F \in \mathcal{D}$, the (pathwise) directional derivative

$$D_\gamma F(\omega) := \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left[F(\omega + \varepsilon\gamma) - F(\omega)\right]$$

can be defined for directions γ not necessarily being Cameron-Martin functions, but even Hölder- or merely continuous functions. We furthermore provide sufficient conditions such that $D_\gamma F(\omega) = \int_I DF(t,\omega)\mathrm{d}\gamma(t)$, where DF is the Malliavin derivative of F. We extend this approach to discontinuous functions as far as regulated functions and define a notion of directional derivative in the direction of a measure. The main results are summarized in Table 3.1.

<table>
<tr><th></th><th>$D_\gamma F = \sum \beta_{pm}\gamma_{pm}$</th><th>$\int_I DF \mathrm{d}\gamma$ exists</th><th>$D_\gamma F = \int_I DF \mathrm{d}\gamma$</th></tr>
<tr><td>$\gamma \in \mathcal{H}$
Section 3.3</td><td>$\gamma_{pm} = \int_I H_{pm}(t)\dot{\gamma}(t)\mathrm{d}t$
Theorem 3.3.2</td><td>Theorem 3.3.4</td><td>Theorem 3.3.5</td></tr>
<tr><td>$\gamma \in C_0^\alpha(I)$
Section 3.4</td><td>Theorem 3.4.15</td><td>Theorem 3.4.18</td><td>Theorem 3.4.19
with (3.4.7)</td></tr>
<tr><td>$\gamma \in C_0(I)$
Section 3.5</td><td colspan="3">Theorem 3.5.4</td></tr>
<tr><td>$\gamma \in BV(I)$
Section 3.6</td><td colspan="3">Theorem 3.6.2 and Theorem 3.6.6</td></tr>
<tr><td>$\gamma \in \mathfrak{R}(I)$
Section 3.6</td><td colspan="2">Theorem 3.6.5</td><td></td></tr>
<tr><td>$\gamma \in \mathfrak{M}(I)$
Section 3.7</td><td>Theorem 3.7.4</td><td>Theorem 3.7.5</td><td>Corollary 3.7.6</td></tr>
</table>

Table 3.1: Overview on main results

A | Appendix to Chapter 2

A.1 The BMO space

For a probability measure $\mathbb{Q}$, recall from Section 2.2 that by $\mathcal{H}_{\text{BMO}}(\mathbb{Q})$ we denote the space of processes $Z \in \mathcal{H}^p(\mathbb{Q})$ for any $p \geq 2$ such that for some constant $K_{BMO} > 0$

$$\sup_{\tau \in \mathcal{T}_{[0,T]}} \mathbb{E}^{\mathbb{Q}} \left[\int_\tau^T |Z_s|^2 \mathrm{d}s \Big| \mathcal{F}_\tau \right] \leq K_{BMO} < \infty,$$

where $\mathcal{T}_{[0,T]}$ is the set of all stopping times $\tau \in [0,T]$. As an easy consequence, if $Z \in \mathcal{H}_{\text{BMO}}(\mathbb{Q})$, then $\int H \mathrm{d}Z \in \mathcal{H}_{\text{BMO}}(\mathbb{Q})$ for any bounded adapted process H.

For more information on BMO spaces and their relation with BSDEs see Subsection 2.3 in [IDR10] or Section 10.1 in [Tou13]; for reference's sake, we state the relevant part (for our purpose) of [IDR10, Lemma 2.2] in the next result.

Lemma A.1.1. *Let $Z \in \mathcal{H}_{BMO}$ and define $\Phi_\cdot := \int_0^\cdot Z_s \mathrm{d}W_s$. Then we have:*

1) *The stochastic exponential $\mathcal{E}(\Phi_T)$ is uniformly integrable.*

2) *There exists a number $r > 1$ such that $\mathcal{E}(\Phi_T) \in L^r$. This property follows from the* Reverse Hölder inequality. *The maximal r with this property can be expressed explicitly in terms of the BMO norm of $\Phi_\cdot$. There exists as well an upper bound for $\|\mathcal{E}(\Phi_T)\|_{L^r}^r$ depending only on T, r and the BMO norm of Φ.*

A.2 Basics of Malliavin's calculus

We briefly introduce the main notation of the stochastic calculus of variations also known as Malliavin's calculus. For more details, we refer the reader to [Nua06], for its application to BSDEs we refer to [Imk08]. Let $\mathcal{S}$ be the space of random variables of the form

$$\xi = F\Big((\int_0^T h_s^{1,i} \mathrm{d}W_s^1)_{1 \leq i \leq n}, \cdots, (\int_0^T h_s^{d,i} \mathrm{d}W_s^d)_{1 \leq i \leq n} \Big),$$

where $F \in C_b^\infty(\mathbb{R}^{n \times d})$, $h^1, \cdots, h^n \in L^2([0,T]; \mathbb{R}^d)$, $n \in \mathbb{N}$. To simplify notation, assume that all h^j are written as row vectors. For $\xi \in \mathcal{S}$, we define $D = (D^1, \cdots, D^d) : \mathcal{S} \to L^2(\Omega \times [0,T])^d$ by

$$D_\theta^i \xi = \sum_{j=1}^n \frac{\partial F}{\partial x_{i,j}} \Big(\int_0^T h_t^1 \mathrm{d}W_t, \ldots, \int_0^T h_t^n \mathrm{d}W_t \Big) h_\theta^{i,j}, \quad 0 \leq \theta \leq T, \quad 1 \leq i \leq d,$$

and for $k \in \mathbb{N}$ its k-fold iteration by $D^{(k)} = (D^{i_1} \cdots D^{i_k})_{1 \le i_1, \cdots, i_k \le d}$. For $k \in \mathbb{N}$, $p \ge 1$ let $\mathbb{D}^{k,p}$ be the closure of $\mathcal{S}$ with respect to the norm

$$\|\xi\|_{k,p}^p = \mathbb{E}\Big[\|\xi\|_{L^p}^p + \sum_{i=1}^k \||D^{(k)}\xi|\|_{(\mathcal{H}^p)^i}^p\Big].$$

$D^{(k)}$ is a closed linear operator on the space $\mathbb{D}^{k,p}$. Observe that if $\xi \in \mathbb{D}^{1,2}$ is $\mathcal{F}_t$-measurable, then $D_\theta \xi = 0$ for $\theta \in (t,T]$. Further denote $\mathbb{D}^{k,\infty} = \cap_{p>1} \mathbb{D}^{k,p}$. We also need Malliavin calculus for $\mathbb{R}^m$-valued smooth stochastic processes. For $k \in \mathbb{N}, p \ge 1$, denote by $\mathbb{L}^{k,p}(\mathbb{R}^m)$ the set of $\mathbb{R}^m$-valued progressively measurable processes $u = (u^1, \cdots, u^m)$ on $[0,T] \times \Omega$ such that

i) for Lebesgue-a.a. $t \in [0,T]$, $u(t,\cdot) \in (\mathbb{D}^{k,p})^m$;

ii) $[0,T] \times \Omega \ni (t,\omega) \mapsto D^{(k)}u(t,\omega) \in (L^2([0,T]^{1+k}))^{d \times n}$ admits a progressively measurable version;

iii) $\|u\|_{k,p}^p = \|u\|_{\mathcal{H}^p}^p + \sum_{i=1}^k \| D^i u \|_{(\mathcal{H}^p)^{1+i}}^p < \infty$.

Note that Jensen's inequality gives[1] for all $p \ge 2$

$$\mathbb{E}\Big[\Big(\int_0^T \int_0^T |D_u X_t|^2 \mathrm{d}u\, \mathrm{d}t\Big)^{\frac{p}{2}}\Big] \le T^{p/2-1} \int_0^T \|D_u X\|_{\mathcal{H}^p}^p \mathrm{d}u.$$

We recall a result from [Imk08] concerning the rule for the Malliavin differentiation of Itô integrals which is of use in applications of Malliavin's calculus to stochastic analysis.

Theorem A.2.1 (Theorem 2.3.4 in [Imk08])**.** *Let $(X_t)_{t\in[0,T]} \in \mathcal{H}^2$ be an adapted process and define $M_t := \int_0^t X_r \mathrm{d}W_r$ for $t \in [0,T]$. Then, $X \in \mathbb{L}^{1,2}$ if and only if $M_t \in \mathbb{D}^{1,2}$ for any $t \in [0,T]$.*
Moreover, for any $0 \le s,t \le T$ we have $D_s M_t = X_s \mathbb{1}_{\{s \le t\}}(s) + \mathbb{1}_{\{s\le t\}}(s) \int_s^t D_s X_r \mathrm{d}W_r$.

A.3 Basic Malliavin calculus results for SDEs

With relation to the Brownian motions W^R and W^S, we denote the Malliavin differential operators D^{W^R} and D^{W^S}, see Appendix A.2.

Proposition A.3.1. *Let Assumption 2.6.1 hold. Then SDEs* (2.3.1) *and* (2.3.2) *have a unique solution $R, S \in \mathcal{S}^p$ for any $p \ge 2$ and*

i) $R, S \in \mathbb{D}^{1,2}$. We have $D_u^{W^S} R_t = D_u^{W^R} S_t = 0$ for any $t,u \in [0,T]$ as well as

$$D_u^{W^R} R_t = \mathbb{1}_{\{u\le t\}} b \qquad \text{and} \qquad D_u^{W^S} S_t = \mathbb{1}_{\{u \le t\}} \sigma^S S_t, \qquad t,u \in [0,T]; \tag{A.3.1}$$

ii) For any jointly measurable function $\psi : [0,T] \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ that is Lipschitz continuous (in the second spatial variable), it holds that

$$D_u^{W^R}\big(\psi(t,S_t,R_t)\big) = D_r^{W^R}\big(\psi(t,S_t,R_t)\big) \qquad \forall u,r \in [0,t], \quad t \in [0,T]. \tag{A.3.2}$$

Furthermore, $\Big(D_0^{W^R}\big(\psi(\cdot, S_\cdot, R_\cdot)\big)\Big) \in \mathcal{S}^\infty$.

[1]The reason behind this last inequality is that within the BSDE framework the usual tools to obtain a priori estimates yield with much difficulty the LHS while with relative ease the RHS.

iii) $H^D, H^a \in \mathbb{L}^{1,2} \cap \mathcal{S}^\infty$ *for any* $a \in \mathbb{A}$ *(recall* (2.6.1)*) and there exists* $M > 0$ *for any* $0 \leq r, u \leq T$ *and any* $\zeta \in \mathbb{A} \cup \{D\}$ *such that* $D_u^{W^R} H^\zeta = D_r^{W^R} H^\zeta$ *and* $0 < |D_\cdot^{W^R} H^\zeta| \leq M$.

iv) Let $\zeta \in \mathbb{A} \cup \{D\}$ *and let* $r_0 \in \mathbb{R}$. *The mapping* $r_0 \mapsto (D_u^{W^R} H^\zeta)$ *is Lipschitz continuous uniformly in* $u \in [0, T]$ *for any* $s_0 \in (0, +\infty)$.

Proof. Throughout let $\zeta \in \mathbb{A} \cup \{D\}$. General results on SDEs follow from e.g. Section 2 in [IDR10][2], standard Malliavin calculus, the fact that S is a Geometric Brownian motion and $\mu^R \in C([0, T]; \mathbb{R})$.

Proof of i) The identity $D_u^{W^S} R_t = D_u^{W^R} S_t = 0$ is trivial.

Proof of ii) We prove (A.3.2): assume ψ to be differentiable, then for $u, r \in [0, t]$

$$D_u^{W^R}\big(\psi(t, S_t, R_t)\big) = (\partial_{x_2}\psi)(t, S_t, R_t) b = D_r^{W^R}\big(\psi(t, S_t, R_t)\big),$$

where we used (A.3.1). Now a standard approximation by mollification delivers the two results.

Proof of iii) The form of the $\mathcal{F}_T$-measurable payoffs H^D, H^a is quite specific and it is clear that for $0 \leq u \leq T$ and $\zeta \in \mathbb{A} \cup \{D\}$

$$D_u^{W^R} H^\zeta = D_u^{W^R}\big(h^\zeta(S_T, R_T)\big) = \big\langle (\nabla h^\zeta)(S_T, R_T), (0, \mathbb{1}_{\{u \leq T\}} b)\big\rangle = b(\partial_{x_2} h^\zeta)(S_T, R_T) \tag{A.3.3}$$

The boundedness of $D_\cdot^{W^R} H^\zeta$ follows from uniform boundedness of the derivatives of $h^\zeta \in C_b^2$. We can then conclude that if $\partial_{x_2} h^\zeta \neq 0$, then it follows that $D_\cdot^{W^R} H^\zeta \neq 0$ and, moreover, the identity $D_u^{W^R} H^\zeta = D_r^{W^R} H^\zeta$ follows from (A.3.2).

Proof of iv) We now close with the proof of the last statement. Take $s_0 \in (0, +\infty)$ and let $r_0, \widetilde{r}_0 \in \mathbb{R}$ be two initial conditions for R (see (2.3.1)) and we denote the corresponding SDE solutions R and $\widetilde{R}$ respectively. We also denote H^ζ and $\widetilde{H}^\zeta$ the random variables depending on R and $\widetilde{R}$ respectively. Due to the linear form of (2.3.1) it is immediate that $R_t - \widetilde{R}_t = r_0 - \widetilde{r}_0$ for any $t \in [0, T]$.

The properties of $|D_u^{W^R} H^\zeta - D_u^{W^R} \widetilde{H}^\zeta|$ follow from those of $\partial_{x_2} h^\zeta$ and (A.3.3). By assumption h^ζ is twice continuously differentiable (in space) with bounded derivatives, hence, for some $K \geq 0$

$$\big|(\partial_{x_2} h^\zeta)(S_T, R_T) - (\partial_{x_2} h^\zeta)(S_T, \widetilde{R}_T)\big| \leq K|R_T - \widetilde{R}_T| = K|r_0 - \widetilde{r}_0|.$$

It follows that for some constant $C \geq 0$ independent of the data u, s_0, r_0 and $\widetilde{r}_0$ one has, as required, $|D_u^{W^R} H^\zeta - D_u^{W^R} \widetilde{H}^\zeta| \leq C|r_0 - \widetilde{r}_0|$. □

A.4 Minor calculations

Lemma A.4.1. *For* $N \in \mathbb{N}$, $N \geq 2$, *we have*

$$\sum_{k=2}^{N} \frac{k-1}{(N-1)^k} \binom{N}{k} = 1.$$

[2] Specifically, Theorem 2.3 in [IDR10] asserts the existence of a unique solution of a sufficiently nice SDE.

Proof. With $k = N - j$, an index change and the binomial theorem one obtains

$$
\begin{aligned}
&\sum_{k=2}^{N} \frac{k-1}{(N-1)^k} \binom{N}{k} \\
&= \sum_{j=0}^{N-2} \frac{N-1}{(N-1)^{N-j}} \binom{N}{j} \\
&= (N-1) \sum_{j=0}^{N-2} \left(\frac{1}{N-1} \right)^{N-j} \binom{N}{j} - \sum_{j=1}^{N-2} \frac{j}{(N-1)^{N-j}} \cdot \frac{N}{j} \binom{N-1}{j-1} \\
&= (N-1) \sum_{j=0}^{N-2} \left(\frac{1}{N-1} \right)^{N-j} \binom{N}{j} - N \sum_{j=0}^{N-3} \left(\frac{1}{N-1} \right)^{N-1-j} \binom{N-1}{j} \\
&= (N-1) \left[\sum_{j=0}^{N} \left(\frac{1}{N-1} \right)^{N-j} \binom{N}{j} - \frac{N}{N-1} - 1 \right] \\
&\quad - N \left[\sum_{j=0}^{N-1} \left(\frac{1}{N-1} \right)^{N-1-j} \binom{N-1}{j} - \frac{N-1}{N-1} - 1 \right] \\
&= (N-1) \left[\left(\frac{N}{N-1} \right)^{N} - \frac{N}{N-1} - 1 \right] - N \left[\left(\frac{N}{N-1} \right)^{N-1} - 2 \right] = 1.
\end{aligned}
$$

□

B Appendix to Chapter 3

B.1 Integration theory

B.1.1 Different types of integration

Even though we are interested in directional derivatives, the main issues arise in the field of integration theory. There are countless books devoted to this subject, for instance [Bur07], [Gor94], [Hil63], [McL80], [vBC00], [Els11], [Nie97], just to name a few. We will briefly introduce the integrals that we use and emphasize how the different types of integrals are related to one another.

For the sake of consistency, we present all integration results on the interval $I = [0, 1]$.

A partition of the interval $I = [0, 1]$ is given by nonoverlapping subintervals $[t_{k-1}, t_k]$ such that $0 = t_0 < t_1 < \ldots < t_n = 1$. The partition is called *tagged* if there are chosen distinct numbers $z_k \in [t_{k-1}, t_k]$ for $k = 1, 2, \ldots, n$, called *tags*. The Riemann integral of a (sufficiently nice) function $f\colon I \to \mathbb{R}$, denoted by $(\mathcal{R}) \int_I f(t)\mathrm{d}t$, is given by the real number A satisfying that for each $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\left| A - \sum_{k=1}^{n} f(z_k)(t_k - t_{x-1}) \right| < \varepsilon$$

for any partition of I satisfying $t_k - t_{k-1} < \delta$ for $k = 1, 2, \ldots, n$; we call such a partition *δ-fine*. That integral, if it exists, is independent of the partition and the chosen tags. There are various extensions of this elementary Riemann integral, allowing for instance the integration of unbounded functions and integration on an unbounded domain.

A rather recent extension is known under several names such as *generalized Riemann integral* or *gauge integral* or *Henstock-Kurzweil integral*, where the latter goes back to the seminal papers of J. Kurzweil (1957) and R. Henstock (1961). The integrals of A. Denjoy (1912) and O. Perron (1914) have been shown to be equivalent, see for instance [Gor94].[1] The idea of the generalized Riemann integral is that the partition need not be equally fine on the entire domain of integration, but instead it needs to be fine when f is less regular and it can be coarse when f is more regular. We introduce the notion of a gauge, i.e., an interval-valued function γ on I which assigns to each point $t \in I$ a neighborhood of that point. For a fixed gauge γ, a tagged partition is called γ-fine if $[t_{k-1}, t_k] \subset \gamma(z_k)$ for $k = 1, \ldots, n$. A function f is called Riemann integrable in the generalized sense if there exists a real number A such that for each $\epsilon > 0$ there

[1] The related integral of McShane is slightly weaker, see for instance [Swa01, Chapter 9] or [Pfe93, Chapters 2 and 6].

exists a gauge γ such that $|A - \sum_{k=1}^{n} f(z_k)(t_k - t_{k-1})| < \varepsilon$ for every tagged partition such that $[t_{k-1}, t_k] \subset \gamma(z_k)$ for $k = 1, 2, \ldots, n$. We denote the gauge integral of f by $(\mathcal{G}) \int_I f(t)\mathrm{d}t$. The advantage of gauge integrals over Lebesgue and Riemann integrals is that functions with antiderivatives are integrable, even without the requirement of boundedness necessary for the Lebesgue fundamental theorem of calculus. For more details on gauge integrals, see [Swa01] or [KS12].

If a function f belongs to $\mathcal{M}(I)$, i.e. it is measurable, then measure-theoretic induction gives a reasonable definition of an integral, namely what is also called (abstract) Lebesgue integral (see e.g. [Els11] on measure theory[2]):

- if $f(t) = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}(t)$ for some sets $A_1, \ldots, A_n \in \mathfrak{B}(I)$ (i.e., f is a simple function), then $\int_I f(t)\mathrm{d}\mu(t) = \sum_{k=1}^{n} a_k \mu(A_k)$;
- if $f \in \mathcal{M}(I)$ is positive, then it can be approximated from below by a sequence of simple functions $(f_n)_{n\in\mathbb{N}}$ and, keeping in mind the continuity from below of measure μ, the integral $\int_I f(t)\mathrm{d}\mu(t)$ is defined as $\lim_{n\to\infty} \int_I f_n(t)\mathrm{d}\mu(t)$,
- if $f \in \mathcal{M}(I)$, we can consider its positive and negative part, for which we know the integral already, and define $\int_I f(t)\mathrm{d}\mu(t) := \int_I f^+(t)\mathrm{d}\mu(t) - \int_I f^-(t)\mathrm{d}\mu(t)$, provided at least one of the two integrals on the RHS is finite.

The Lebesgue integral includes the basic Riemann integral, but not the improper Riemann integral. One of its biggest advantages is the availability of powerful convergence results, which do not hold for the basic Riemann integral. The generalized Riemann integral, however, includes both the Lebesgue integral and the improper Riemann integral and it features those convergence results. If we denote by $\mathcal{R}(I)$, $\mathcal{L}(I)$ and $\mathcal{G}(I)$ the spaces of Riemann-, Lebesgue- and gauge-integrable functions on I, respectively, then $\mathcal{R}(I) \subset \mathcal{L}(I) \subset \mathcal{G}(I)$.

Example B.1.1. *Let $f\colon I \to \mathbb{R}$ be given by*

$$f(t) = \begin{cases} t^2 \cos(\frac{1}{t^2}) & ,\ t \in (0,1], \\ 0 & ,\ t = 0. \end{cases}$$

The improper Riemann integral (and also the generalized Riemann integral)

$$\int_I f'(t)\mathrm{d}t = \lim_{a\to 0} \int_a^1 f'(t)\mathrm{d}t = f(1) = \cos(1)$$

exists, whereas $f' \notin \mathcal{L}(I)$. The reason is that Lebesgue integrable functions are by definition absolutely integrable and $\int_I |f'(t)|\mathrm{d}t = \infty$, hence f' cannot be Lebesgue integrable on this interval. Another example would be the "sawtooth function" presented in [Bur07, Section 8.7.5] *or the piecewise constant function presented in* [Bar01, Example 2.8].

Example B.1.2. *The indicator function $f(t) = \mathbb{1}_{\mathbb{Q}\cap I}(t)$ on I is known to be Lebesgue-, but not Riemann-integrable on I. It is Riemann integrable in the generalized sense with*

[2] Elstrodt [Els11, IV §3 Section 6] mentions that this approach to the Lebesgue integral via monotone approximating sequences stems originally from W. H. Young and was the key to the LS integral.

$(\mathcal{G})\int_I \mathbb{1}_{\mathbb{Q}}(t)\mathrm{d}t = 0$ (cf. [McL80, Example 4 in Section 1.2]*). To see this, let $\mathbb{Q}\cap I = \{q_n \mid n\in\mathbb{N}\}$. For a given $\varepsilon > 0$ choose the gauge γ as follows:*

$$\gamma(t) := \begin{cases} (t-\frac{\varepsilon}{2^{n+2}}, t+\frac{\varepsilon}{2^{n+2}}) & , t = q_n \in \mathbb{Q}, \\ (t-1, t+1) & , t \in I\setminus\mathbb{Q}, \end{cases}$$

where γ associates to every point in I an open interval in $\mathbb{R}$, which need not be contained in I. The worst possible choice of tags would be $z_n = q_n$ (modulo reordering) and the corresponding Riemann sum is then

$$\sum_{n=1}^{N} f(q_n)(t_{n+1}-t_n) = \sum_{n=1}^{N} 1\cdot(t_{n+1}-t_n) \leq \sum_{n=1}^{N}\frac{\varepsilon}{2^{n+1}}$$

Each rational q_n can be at most twice a tag (as left and right endpoint of adjacent intervals), hence for $N\to\infty$ this sum is still smaller than ε, which proves that the value of the generalized Riemann integral $(\mathcal{G})\int_I \mathbb{1}_{\mathbb{Q}}(t)\mathrm{d}t$ is zero.

The next extension is that towards Stieltjes integrals of the different types. The RS integral $(\mathcal{R})\int_I f(t)\mathrm{d}g(t)$ for a non-increasing function g is approximated by Riemann sums of the structure $\sum_{k=1}^{n} f(z_k)\,[g(t_k)-g(t_{k-1})]$ for δ-fine partitions, i.e., for $\delta > 0$, $t_k - t_{k-1} < \delta$. The generalized RS integral $(\mathcal{G})\int_I f(t)\mathrm{d}g(t)$ has the same Riemann sums, but the tagged partition is called γ-fine if for a gauge γ, $[t_{k-1},t_k]\subset\gamma(z_k)$. For more details, see e.g. [McL80].

The LS integral is again defined with the help of the LS measure. For a monotone function g on I one can define $\mu_g((a,b]) := g(b_+) - g(a_+)$, $\mu_g([a,b]) := g(b_+) - g(a_-)$ and so on for $a,b\in I$ with $a<b$ (see [vBC00, Chapter 4]). Thus one obtains a measure μ_g on I. The LS integral $(\mathcal{L})\int_I f(t)\mathrm{d}g(t)$ is then defined as the Lebesgue integral $\int_I f(t)\mathrm{d}\mu_g(t)$. The Stieltjes integrals include the non-Stieltjes integrals of the corresponding types through the choice $g(t) = t$.

Young's Stieltjes integral [You36] is of RS type. While the RS integral was known to exist whenever the integrand is continuous and the integrator of bounded variation or the other way round, Young used Hölder's inequality to allow for functions of bounded p- and q-variation, respectively, provided $\frac{1}{p}+\frac{1}{q} > 1$.

A slightly stricter version, namely Hölder continuous integrand and integrator with Hölder coefficients α and β, respectively, permits an extension beyond $\alpha+\beta > 1$ by means of *paracontrol*, see for instance [Gub04] or [GIP16]. It should be mentioned that while the YS integral permits discontinuities of integrand and integrator, the concept of paracontrol cannot be applied to discontinuous functions.

We only work on the interval I, hence the notation $f\in\mathcal{R}(g)$ (or $f\in\mathcal{L}(g)$ or $f\in\mathcal{G}(g)$) signifies that the function f is Riemann- (resp. Lebesgue- resp. generalized Riemann-) Stieltjes integrable on I with respect to a function g. As is shown in Section 8.5 in [Nie97], if $g\in C(I)$, then $\mathcal{R}(g)\subset\mathcal{L}(g)$. It is also mentioned that this is not true in general if the continuity is dropped, i.e., there exist functions with a finite improper RS integral which are not LS integrable. Both LS and improper RS integrals are included in the Stieltjes version of the *gauge integrals* or *generalized Riemann integrals*, providing maximum liberty.

Further relations and inclusions between different Stieltjes integrals are presented in [DN11, Figure 2.1]. While the YS integral can go beyond the LS integral, the relation between YS and generalized RS integration is not clear. These competing notions of Stieltjes integral therefore do not appear simultaneously; the former appears only in Section 3.4 and the latter in Section 3.6.

Example B.1.3. *If a set $A \subset \mathbb{R}$ is not Lebesgue measurable, then $\mathbb{1}_A$ cannot be LS integrable against any function g, even the constant function. However, for $g \equiv c$ for some $c \in \mathbb{R}$ one has*

$$(\mathcal{R}) \int_{\mathbb{R}} \mathbb{1}_A(t) \mathrm{d}g(t) = (\mathcal{G}) \int_{\mathbb{R}} \mathbb{1}_A(t) \mathrm{d}g(t) = 0.$$

Along the lines of Example B.1.1 one can construct examples of functions $f \in \mathcal{G}(g)$ with $f \notin \mathcal{L}(g)$. Take for instance $g(t) = t - a$ for some $a \in \mathbb{R}$ and f' as in Example B.1.1 and consider the integral $\int_I f'(t) \mathrm{d}g(t)$. On an unbounded domain one can alternatively consider $g(t) = t - a$ and $f(t) = \frac{\sin(t)}{t}$ for $t \in (0, \infty)$. Recall that $(\mathcal{R}) \int_0^\infty f(t) \mathrm{d}g(t) = (\mathcal{R}) \int_0^\infty \frac{\sin(t)}{t} \mathrm{d}t = \frac{\pi}{2}$, yet $f \notin \mathcal{L}((0, \infty))$ and it is not LS integrable w.r.t. g.

Remark B.1.4. *Any function that is Riemann integrable in the generalized sense, i.e., which belongs to $\mathcal{G}(I)$, is also measurable. See e.g.* [Bar01, Theorem 6.8] *or* [Swa01, Theorem 3 in Appendix 3]. *Thus a non-measurable function like in the above example cannot be Riemann integrable in the generalized sense.*

B.1.2 Available convergence results

One rather strict, but useful convergence result is that on uniform convergence. A criterion for uniform convergence of a series is the Weierstrass M-test, see Theorem B.2.8. A convergence theorem for Young's Stieltjes integrals relying on uniform convergence is presented in [You36] (called *Theorem on term by term integration*).

As to the dominated convergence theorem known from Lebesgue integration, there is a (stricter) version for RS integrals. It is found in the literature under the names of *bounded convergence theorem* or *Arzelà-Osgood theorem*, see for instance [Hil63, Theorem II.15.6 or II.15.9] or [Wes51] for different versions of this theorem.

For LS integration, the situation is much easier and, as long as the integrator γ is at least continuous, $\mathcal{R}(\gamma) \subset \mathcal{L}(\gamma)$, which permits us to apply the Stieltjes versions of the well-known results of monotone convergence, dominated convergence and Fatou's Lemma. They can be found for instance in [Nie97, Section 8.4].

Collection of results

For all of the following results, recall that we fix $I = [0, 1]$.

Theorem B.1.5 ([Bur07, Theorem 4.2.1])**.** *If $f \in C(I)$ and g is a step function on I, then $f \in \mathcal{R}(g)$.*

Theorem B.1.6 ([Bur07, Theorem 4.3.2])**.** *Let $J = [a, b] \subset I$ denote a fixed closed subinterval of I. If f and g are bounded functions on J with no common discontinuities and if $f \in \mathcal{R}(g)$, then $g \in \mathcal{R}(f)$ and we have the integration by parts formula*

$$\int_a^b f(t) \mathrm{d}g(t) = f(b)g(b) - f(a)g(a) - \int_a^b g(t) \mathrm{d}f(t).$$

Theorem B.1.7 ([McL80, page 199]). *Let $f, g\colon I \to \mathbb{R}$ with one of them being of bounded variation and the other regulated. Then both $(\mathcal{G}) \int_I f \mathrm{d}g$ and $(\mathcal{G}) \int_I g \mathrm{d}f$ exist.*

Theorem B.1.8 (Uniform convergence of gauge integral, [Bar01, Theorem 8.3]). *If $f_n \in \mathcal{G}(I)$ for $n \in \mathbb{N}$ and if f_n converges to f uniformly on I, then $f \in \mathcal{G}(I)$ and $(\mathcal{G}) \int_I f(t)\mathrm{d}t = \lim_{n\to\infty}(\mathcal{G}) \int_I f_n(t)\mathrm{d}t$.*

Theorem B.1.9 (Uniform convergence of the generalized RS integral, [Sch96, Theorem 24.17]). *Let $f_n\colon I \to \mathbb{R}$ ($n \in \mathbb{N}$) and $f\colon I \to \mathbb{R}$ be such that $f_n \to f$ uniformly on I for $n \to \infty$. If $g\colon I \to \mathbb{R}$ is increasing and $f_n \in \mathcal{G}(g)$ for $n \in \mathbb{N}$, then $f \in \mathcal{G}(g)$ and $\lim_{n\to\infty} \int_I f_n(t)\mathrm{d}g(t) = \int_I f(t)\mathrm{d}g(t)$.*

This result can be extended without any changes to $g \in BV(I)$, see [McL80, page 194]. We also have the corresponding result on finite measure spaces:

Theorem B.1.10 (Uniform convergence in finite measure space). *Let μ be a finite measure on $(I, \mathfrak{B}(I))$ and let $(f_n)_{n\in\mathbb{N}}$ be a sequence of bounded measurable functions on I converging uniformly to f. Then $\lim_{n\to\infty} \int_I f_n(t)\mathrm{d}\mu(t) = \int_I f(t)\mathrm{d}\mu(t)$.*

This can be easily seen from the inequality $\int_I |f_n(t) - f(t)| \,\mathrm{d}\mu(t) \leq \mu(I) \|f_n - f\|_\infty$ for any $n \in \mathbb{N}$.

Theorem B.1.11 (Dominated convergence of the abstract Lebesgue integral, [Bur07, Theorem 6.3.3] or [Gor94, Theorem 3.25/Corollary 13.5]). *For $n \in \mathbb{N}$ let $f_n \in \mathcal{L}(I)$ (Lebesgue-integrable on I) converging a.e. to f on I. Let $g \in \mathcal{L}(I)$ s.t. $|f_n| \leq g$ for all $n \in \mathbb{N}$ on I. Then $f \in \mathcal{L}(I)$ and $\int_I f(t)\mathrm{d}\mu(t) = \lim_{n\to\infty} \int_I f_n(t)\mathrm{d}\mu(t)$.*

Theorem B.1.12 (Dominated convergence of LS integral, [Nie97, Theorem 8.11]). *Let u be nondecreasing on I and $f_n \in \mathcal{L}(u)$ for $n \in \mathbb{N}$ a sequence converging pointwise to f such that $|f_n| \leq g$ for all $n \in \mathbb{N}$ and some $g \in \mathcal{L}(u)$. Then $f \in \mathcal{L}(u)$ and $(\mathcal{L}) \int_I f(t)\mathrm{d}u(t) = \lim_{n\to\infty}(\mathcal{L}) \int_I f_n(t)\mathrm{d}u(t)$.*

Theorem B.1.13 (Vitali convergence theorem, [Gor94, Theorem 13.3]). *For $n \in \mathbb{N}$ let $f_n \in \mathcal{L}(I)$ and $F_n(t) := \int_0^t f_n(s)\mathrm{d}\mu(s)$ ($t \in I$). Suppose that f_n converges pointwise to f on I. If $\{F_n\}_{n\in\mathbb{N}}$ is equi-absolutely continuous[3] on I, then $f \in \mathcal{L}(I)$ and $\int_I f(t)\mathrm{d}\mu(t) = \lim_{n\to\infty} \int_I f_n(t)\mathrm{d}\mu(t)$.*

B.2 Relevant spaces and collection of related results

B.2.1 The space $C([0,1])$ as inner product space

Let $I = [0,1]$. Obviously $C(I) \subset L^2(I)$ (as continuous functions on a compact set are bounded). Therefore, $\langle f, g\rangle := \langle f, g\rangle_{L^2(I)} = \int_0^1 f(t)g(t)\mathrm{d}t$ is an inner product on $C(I)$. The norm induced by $\langle \cdot, \cdot\rangle$ is $\|f\|_{L^2(I)} = \sqrt{\langle f, f\rangle} = \left(\int_I f^2(t)\mathrm{d}t\right)^{1/2}$.
The supremum norm $\|f\|_\infty := \sup\{|f(t)| \mid t \in I\}$ is not induced by any inner product, for it does not satisfy the parallelogram identity $\|f+g\|^2 + \|f-g\|^2 = 2\left(\|f\|^2 + \|g\|^2\right)$. This can be easily seen by plugging in $f(t) = t$ and $g(t) = 1 - t$, both defined on I.
The space $C(I)$, equipped with the L^2-inner product and norm, is not complete; hence it is not a Hilbert space.

[3] $\{\int_\cdot f_n\}$ is equi-absolutely continuous w.r.t. the Lebesgue-measure $\mu|_I$ if for each $\varepsilon > 0$ there ex. $\delta > 0$ s.t. $\int_J |f_n| < \varepsilon$ for all $n \in \mathbb{N}$ whenever $J \subset I$ measurable with $\mu(J) < \delta$.

Theorem B.2.1 ([Haa10, Chapter III Section §2]). *Let $f \in C(I)$ and for $(p,m) \in \Lambda_0$ define $a_{pm} := \int_I f(t)H_{pm}(t)\mathrm{d}t$. Then $\sum_{(p,m)\in\Lambda_0} a_{pm}H_{pm}(t)$ converges to f uniformly on I.*

B.2.2 Step functions and regulated functions

We collect here the definitions of step and regulated functions and some essential properties. The results stem from [Lea01, §4.4].

Definition B.2.2 (Step function). *We call $f\colon I \to \mathbb{R}$ a step function on I if there exists a finite number of disjoint (possibly degenerate) intervals $I_k \subset I$, $k = 1,\dots,n$, such that $I = \bigcup_{k=1}^n I_k$ and f is constant on each I_k.*

According to [Lea01, Theorem 1] this is equivalent to f having only finitely many points of discontinuity and $f(I)$ being a finite set.

Definition B.2.3 (Regulated function). *A function $f\colon I \to \mathbb{R}$ is called regulated, if it has left and right limits everywhere, i.e. $f(0_+)$, $f(1_-)$ exist and for $0 < t < 1$, both $f(t_-)$ and $f(t_+)$ exist.*

Examples of regulated functions on I are step functions, functions of bounded variation and continuous functions on I.

Theorem B.2.4 ([Lea01, §4.4 Theorem 3]). *A function $f\colon I \to \mathbb{R}$ is regulated if and only if it is a uniform limit of step functions on I.*

Every regulated function on I is bounded and the product of two regulated functions is again regulated. From the Riemann integrability of step functions and the uniform convergence theorem, regulated functions are Riemann integrable and due to boundedness also Lebesgue integrable.

B.2.3 Collection of necessary results from sequence spaces

Recall the Hölder inequality: $\|xy\|_{\ell^1} \leq \|x\|_{\ell^p} \cdot \|y\|_{\ell^q}$ if $\frac{1}{p} + \frac{1}{p} = 1$. Furthermore, for $1 < p < \infty$ and any given sequence x, $\|x\|_{\ell^\infty} \leq \|x\|_{\ell^p} \leq \|x\|_{\ell^1}$, hence $\ell^1 \subset \ell^p \subset \ell^\infty$.

Definition B.2.5 (Unconditional convergence). *A series $\sum_{n=1}^\infty x_n$ in a Banach space X is called* unconditionally convergent *if the reordered series $\sum_n x_{\pi(n)}$ converges for every permutation π of $\mathbb{N}$.*

Lemma B.2.6. *A series $\sum_{n=1}^\infty x_n$ in a Banach space X is unconditionally convergent if and only if $\sum_{n=1}^\infty \alpha_n x_n$ converges unconditionally for all bounded sequences $(\alpha_n)_n \in \ell^\infty$.*

Lemma B.2.7. *Let $(\alpha_n)_n$ and $(\beta_n)_n$ be $\mathbb{R}$-valued sequences.*

a) If $(\alpha_n)_n \in \ell^p$ and $(\beta_n)_n \in \ell^q$, then $(\alpha_n\beta_n)_n \in \ell^1$, whenever $\frac{1}{p} + \frac{1}{p} = 1$.

b) If $(\alpha_n)_n \in \ell^\infty$ and $(\beta_n)_n \in \ell^1$, then $\sum_{n=1}^\infty \alpha_n\beta_n$ is an unconditionally convergent series.

c) If $(\alpha_n)_n \in \ell^1$ and $(\beta_n)_n \in \ell^\infty$, then $\sum_{n=1}^\infty \alpha_n\beta_n$ is an unconditionally convergent series.

Proof. The first follows from the Hölder inequality, the second and third follow from Lemma B.2.6. Convergence follows from absolute convergence by completeness of the underlying space, because in a complete space absolute convergence implies convergence. □

Theorem B.2.8 (Weierstrass M-test, [Rud76, Theorem 7.10]). *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of functions on I and let $(M_n)_{n\in\mathbb{N}}$ be a sequence of real numbers such that $|f_n(t)| \leq M_n$ for all $n \in \mathbb{N}$ and all $t \in I$. If $\sum_{n=1}^{\infty} M_n < \infty$, then $\sum_{n=1}^{\infty} f_n(t)$ converges uniformly on I.*

Corollary B.2.9. *If the sequence $M_N := \sum_{(p,m)\in\Lambda_0^N} \sqrt{2^p}\,|\beta_{pm}|$ ($N \in \mathbb{N}$) converges, then $\sum_{(p,m)\in\Lambda_0} \beta_{pm} H_{pm}(t)$ defines a uniformly convergent series and is therefore regulated.*

Proof. Recall that $|H_{pm}| \leq \sqrt{2^p}$ for all $(p,m) \in \Lambda_0$ and apply the Weierstrass M-test (Theorem B.2.8) to the sequence of functions $(\sum_{(p,m)\in\Lambda_0^N} \beta_{pm} H_{pm}(t))_{N\in\mathbb{N}}$ to obtain uniform convergence. As the uniform limit of step functions, $\sum_{(p,m)\in\Lambda_0} \beta_{pm} H_{pm}(t)$ is regulated. □

B.2.4 p-variation spaces

Let $V_p(f;I)$ denote the p-variation of a function f on the interval I with values in $\mathbb{R}$, $p \geq 1$. If $\mathfrak{T} = \{0 = t_0 < t_1 < \ldots < t_n = 1 \mid n \in \mathbb{N}\}$ denotes all finite partitions of I, then the p-variation of f on I is

$$V_p(f;I) := \sup_{\mathfrak{T}} \left\{ \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})|^p \right\}.$$

As we only work on the fixed interval I, we simply write $V_p(f)$. We denote by $\mathcal{V}_p(I)$ the space of functions $f\colon I \to \mathbb{R}$ with $V_p(f) < \infty$. For any $p \geq 1$ and $f \in \mathcal{V}_p(I)$, one has $V_p(f) \geq V_p(|f|)$. The following result can be found in [CG98, Remark 2.5]:

Lemma B.2.10. *For $1 \leq p \leq q$, $\mathcal{V}_1(I) \subset \mathcal{V}_p(I) \subset \mathcal{V}_q(I)$.*

A useful property of p-variation, to be found for example as Property (P7) in [CG98], is lower semi-continuity, i.e.:

Lemma B.2.11. *If the sequence of maps $(f_n)_{n\in\mathbb{N}}$ (all mappings from I to $\mathbb{R}$) converges pointwise to $f\colon I \to \mathbb{R}$ as $n \to \infty$, then $V_p(f,I) \leq \liminf_{n\to\infty} V_p(f_n, I)$.*

We also recall here [CG98, Theorem 3.1] for the precise link between Hölder continuity and bounded p-variation.

Theorem B.2.12. *$f\colon I \to \mathbb{R}$ is of bounded p-variation if and only if there exist functions $\phi\colon I \to \mathbb{R}$ and $g\colon \phi(I) \to \mathbb{R}$ such that*

- *ϕ is bounded and nondecreasing,*
- *g is Hölder continuous with exponent $\frac{1}{p}$ and Hölder constant $H(g) \leq 1$,*
- *$f = g \circ \phi$ on I.*

Remark B.2.13. *In rough paths literature (e.g.* [FV10]*), this result usually does not surface, because the object of interest is an integral of continuous integrand and integrator. Hence instead of functions of bounded p-variation the focus lies on continuous functions with bounded p-variation.*

One must be careful not to infer from this theorem that continuity and bounded (p)-variation would already imply $(1/p)-$Hölder continuity, as is highlighted by the following Lemma.

Lemma B.2.14. *If for $I = [0,1]$, $BV(I)$ denotes the set of functions of bounded variation on I and $C(I)$ the set of continuous functions on I, then*

$$(BV(I) \cap C(I)) \setminus \left(\bigcup_{0<\alpha<1} C^\alpha(I) \right) \neq \emptyset.$$

Proof. Consider $f\colon I \to \mathbb{R}$ defined by

$$f(t) := \begin{cases} \frac{1}{\ln(\frac{2}{t})} & \text{for } 0 < t \leq 1, \\ 0 & \text{for } t = 0. \end{cases}$$

By L'Hospital's rule f is continuous at zero and hence on the entire closed interval I. Being monotonically increasing, we know that $f \in BV(I)$. (A function is of bounded variation if and only if it can be written as difference of two monotone functions.) Furthermore, for any $\alpha > 0$

$$\lim_{t\to 0+} \frac{|f(t) - f(0)|}{t^\alpha} = \lim_{t \to 0+} \frac{1}{t^\alpha \ln \frac{2}{t}} = \infty,$$

hence $f \notin C^\alpha(I)$ for any $\alpha > 0$. □

Thus, we see immediately (by letting ϕ be the identity function) that any α-Hölder continuous function has finite $\frac{1}{\alpha}$-variation.

Finally, the following result, which is due to Camille Jordan, is very useful:

Theorem B.2.15 ([Nie97, Theorem 5.10] – Jordan Decomposition Theorem)**.** *A function defined on a closed interval is of bounded variation if and only if it is a difference of two nondecreasing functions.*

B.3 Minor calculations

Lemma B.3.1. *For $m \in \mathbb{N}$ and $\alpha \in (0,1)$ let $M(m,\alpha) := 2(2m-1)^\alpha - (2m-2)^\alpha - (2m)^\alpha$. We have:*

(a) $M(m,\alpha) \geq 0$ for all $m \in \mathbb{N}$ and $\alpha \in (0,1)$.

(b) $M(m,\alpha) \to 0$ as $m \to \infty$ for any $\alpha \in (0,1)$.

Proof. Let's first look at (a): For any $\alpha \in (0,1)$, the function $x \mapsto x^\alpha$ is strictly concave on $\mathbb{R}_+$, i.e. for $x, y \in \mathbb{R}_+$ and $\lambda \in [0,1]$ one has $(\lambda x + (1-\lambda)y)^\alpha \geq \lambda x^\alpha + (1-\lambda)y^\alpha$. Plugging in $x = m-1$ and $y = m$ and $\lambda = \frac{1}{2}$ we get after rearranging terms $2^{1-\alpha}(2m-1)^\alpha \geq (m-1)^\alpha + m^\alpha$, which is again equivalent to $M(m,\alpha) \geq 0$.

For (b), consider $h_\alpha \colon [1,\infty) \to \mathbb{R}$ given by $h_\alpha(m) := m^\alpha - (m-1)^\alpha$ for fixed $\alpha \in (0,1)$. It is a concave mapping, hence for any x, y in its domain, $h_\alpha(y) - h_\alpha(x) \leq h'_\alpha(x)(y-x)$. Specifically for $x = m-1$ and $y = m$ ($m \in \mathbb{N}$) we obtain

$$h_\alpha(m) - h_\alpha(m-1) \leq \alpha m^{\alpha-1} = \frac{\alpha}{m^{1-\alpha}} \overset{m\to\infty}{\longrightarrow} 0.$$

Now $M(m,\alpha) = [h_\alpha(2m-1) - h_\alpha(2m-2)] - [h_\alpha(2m) - h_\alpha(2m-1)]$, which converges to zero as $m \to \infty$ because it is the difference of two sequences, both of which converge to zero. □

Lemma B.3.2. *For $N \in \mathbb{N}_0$ let*

$$t_N := \sum_{n=0}^{N} \left(\frac{-1}{2}\right)^n \quad \text{and} \quad \mathcal{G}_N := \sum_{p=0}^{N} \sqrt{2^p} G_{p1}(t_N).$$

Then

(i) $\mathcal{G}_N = \sum_{p=0}^{\infty} \sqrt{2^p} G_{p1}(t_N)$;

(ii) $(\mathcal{G}_N)_{N\in\mathbb{N}}$ *is strictly increasing;*

(iii) $(\mathcal{G}_N)_{N\in\mathbb{N}}$ *is unbounded.*

Proof. Let us formally introduce $t_{-1} = 0$.

(i) It suffices to observe that for $p > N$, $G_{p1}(t_N) = 0$.

(ii) We first calculate that for $p \in \mathbb{N}_0$, $\sqrt{2^p} G_{p1}(t_N) = (-1)^p\, 2^p\, (t_N - t_{p-1})$. With this,

$$\begin{aligned}
\mathcal{G}_{N+1} - \mathcal{G}_N &= \sum_{p=0}^{N+1} (-2)^p (t_{N+1} - t_{p-1}) - \sum_{p=0}^{N} (-2)^p (t_N - t_{p-1}) \\
&= (t_{N+1} - t_N) \sum_{p=0}^{N+1} (-1)^p\, 2^p \\
&= \left[(-1)^{N+1} \left(\frac{1}{2}\right)^{N+1}\right] \left[\sum_{p=0}^{N+1} (-1)^p\, 2^p\right],
\end{aligned}$$

where both brackets are positive if and only if N is odd and negative if and only if N is even, hence the whole expression is strictly positive for any $N \in \mathbb{N}$.

(iii) From the previous point we easily see that

$$|\mathcal{G}_{N+1} - \mathcal{G}_N| = \left|\sum_{p=0}^{N+1} (-1)^p 2^{p-(N+1)}\right| = \left|\sum_{p=0}^{N+1} (-1)^p \left(\frac{1}{2}\right)^p\right| = |t_{N+1}|,$$

which converges to $\frac{2}{3}$ as $N \to \infty$. If the monotone sequence $(\mathcal{G}_N)_{N\in\mathbb{N}}$ was to be bounded, its increments would have to be a null sequence, which we have just proven to not be the case. This completes the proof.

□

List of Abbreviations

BSDE Backward Stochastic Differential Equation

CN Crossing Network

DM Dealer Market

DP Dark Pool

EMPeR Equilibrium Market Price of external Risk

EMPR Equilibrium Market Price of Risk

FOC first order condition

LHS left-hand side

LS Lebesgue-Stieltjes

LOB limit order book

MPR Market Price of Risk

NE Nash Equilibrium

ONB orthonormal basis

PDE partial differential equation

RHS right-hand side

RS Riemann-Stieltjes

SDE stochastic differential equation

w.r.t. with respect to

YS Young-Stieltjes

List of Figures

List of Tables

Bibliography

[Abe90] A. B. Abel, *Asset Prices under Habit Formation and Catching up with the Joneses*, The American Economic Review **80** (1990), no. 2, 38–42.

[Abe99] A. B. Abel, *Risk premia and term premia in general equilibrium*, Journal of Monetary Economics **43** (1999), no. 1, 3–33.

[ADEH99] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, *Coherent measures of risk*, Math. Finance **9** (1999), no. 3, 203–228. MR1850791 (2002d:91056)

[AIdR10] S. Ankirchner, P. Imkeller, and G. dos Reis, *Pricing and hedging of derivatives based on non-tradable underlyings*, Mathematical Finance **20** (2010April), no. 2, 289–312. MR2354164 (2008h:60220)

[AR08] R. M. Anderson and R. C. Raimondo, *Equilibrium in continuous-time financial markets: Endogenously dynamically complete markets*, Econometrica **76** (2008), no. 4, 841–907.

[Ath13] K. B. Athreya, *Big Ideas in Macroeconomics: A Nontechnical View*, MIT Press, 2013.

[AV15] N. Apergis and D. Voliotis, *Spillover effects between lit and dark stock markets: Evidence from a panel of London Stock Exchange transactions*, International Review of Financial Analysis **41** (2015), 101–106.

[Bar01] R. G. Bartle, *A Modern Theory of Integration*, Graduate Studies in Mathematics, vol. 32, American Mathematical Society, 2001.

[Bat97] R. H. Battalio, *Third Market Broker-Dealers: Cost Competitors or Cream Skimmers?*, The Journal of Finance **52** (1997), no. 1, 341–352.

[BCC16] V. Bally, L. Caramellino, and R. Cont, *Stochastic Integration by Parts and Functional Itô Calculus*, Birkhäuser, 2016.

[BE05] P. Barrieu and N. El Karoui, *Inf-convolution of risk measures and optimal risk transfer*, Finance Stoch. **9** (2005), no. 2, 269–298. MR2211128 (2006k:60010)

[BE09] P. Barrieu and N. El Karoui, *Pricing, Hedging, and Designing Derivatives with Risk Measures*, In R. Carmona (ed.), Indifference Pricing: Theory and Applications., 2009, pp. 77–146.

[Bel06] D. R. Bell, *The Malliavin Calculus*, Dover Publications, Mineola, New York, 2006.

[BHMB16] J. Bielagk, U. Horst, and S. Moreno-Bromberg, *A Principal-Agent Model of Trading Under Market Impact – Crossing networks interacting with dealer markets*, arXiv preprint arXiv:1607.04047 (2016).

[BLDR17] J. Bielagk, A. Lionnet, and G. Dos Reis, *Equilibrium pricing under relative performance concerns*, arXiv preprint arXiv:1511.04218 (2017). Accepted by SIAM Journal on Financial Mathematics.

[BMR00] B. Biais, D. Martimort, and J.-C. Rochet, *Competing mechanisms in a common value environment*, Econometrica **68** (2000July), no. 4, 799–837.

[BRW16] S. Buti, B. Rindi, and I. M. Werner, *Dark Pool Trading Strategies, Market Quality and Welfare*, Journal of Financial Economics (2016). accepted.

[BT04] B. Bouchard and N. Touzi, *Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations*, Stochastic Processes and their Applications **111** (2004), no. 2, 175–206.

[Bur07] F. Burk, *A Garden of Integrals*, The Dolciani Mathematical Expositions, vol. 31, MAA, 2007.

[CEP10] S. N. Cohen, R. J. Elliott, and C. E. M. Pearce, *A General Comparison Theorem for Backward Stochastic Differential Equations*, Advances in Applied Probability **42** (2010), no. 3, 878–898.

[CF10] R. Cont and D.-A. Fournié, *Change of variable formulas for non-anticipative functionals on path space*, Journal of Functional Analysis **259** (2010), no. 4, 1043–1072.

[CF13] R. Cont and D.-A. Fournié, *Functional Itô calculus and stochastic integral representation of martingales*, The Annals of Probability **41** (2013), no. 1, 109–133.

[CG98] V. Chistyakov and O. Galkin, *On maps of bounded p-variation with $p > 1$*, Positivity **2** (1998), no. 1, 19–45.

[CHKP16] P. Cheridito, U. Horst, M. Kupper, and T. A. Pirvu, *Equilibrium pricing in incomplete markets under translation invariant preferences*, Mathematics of Operations Research **41** (2016), no. 1, 174–195.

[Cie59] Z. Ciesielski, *On Haar Functions and on the Schauder Basis of the Space C(0,1)*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys **7** (1959), no. 4, 227–232.

[Cie60] Z. Ciesielski, *On the Isomorphisms of the Spaces H_α and m*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys **8** (1960), no. 4, 217–222.

[CK02] Y. L. Chan and L. Kogan, *Catching Up with the Joneses: Heterogeneous Preferences and the Dynamics of Asset Prices* **110** (2002), no. 6, 1255–1285.

[CK04] M. Capiński and E. Kopp, *Measure, Integral and Probability*, 2nd ed., Springer Undergraduate Mathematics Series, Springer, 2004.

[CLR01] G. Carlier and T. Lachand-Robert, *Regularity of Solutions for Some Variational Problems Subject to a Convexity Constraint*, Communications on Pure and Applied Mathematics **54** (2001), no. 5, 583–594.

[CML16] F. Corsi, S. Marmi, and F. Lillo, *When Micro Prudence Increases Macro Risk: The Destabilizing Effects of Financial Innovation, Leverage, and Diversification*, Operations Research **64** (2016), no. 5, 1073–1088.

[CMV09] F. Caccioli, M. Marsili, and P. Vivo, *Eroding market stability by proliferation of financial instruments*, The European Physical Journal B **71** (2009), no. 4, 467–479.

[CN15] P. Cheridito and K. Nam, *Multidimensional quadratic and subquadratic BSDEs with special structure*, Stochastics **87** (2015), no. 5, 871–884.

[DDH13] T. R. Daniëls, J. Dönges, and F. Heinemann, *Crossing network versus dealer market: Unique equilibrium in the allocation of order flow*, European Economic Review **62** (2013), 41–57.

[DdJK15] H. Degryse, F. de Jong, and V. v. Kervel, *The Impact of Dark Trading and Visible Fragmentation on Market Quality*, Review of Finance **19** (2015), no. 4, 1587–1622, available at `/oup/backfile/Content_public/Journal/rof/19/4/10.1093/rof/rfu027/2/rfu027.pdf`.

[DN11] R. M Dudley and R. Norvaiša, *Concrete Functional Calculus*, Springer Monographs in Mathematics, Springer, 2011.

[dR11] G. dos Reis, *Some advances on quadratic BSDE: Theory - Numerics - Applications*, LAP LAMBERT Academic Publishing, 2011.

[Due49] J. S. Duesenberry, *Income, saving and the theory of consumer behavior*, Harward University Press, 1949.

[Dup09] B. Dupire, *Functional Itô Calculus*, Bloomberg Portfolio Research paper **2009-04** (2009).

[DVAW05] H. Degryse, M. Van Achter, and G. Wuyts, *Crossing Networks: Theory and Evidence*, Tijdschrift voor Bank-en Financiewezen **69** (2005), 114–118.

[DVAW09] H. Degryse, M. Van Achter, and G. Wuyts, *Dynamic order submission strategies with competition between a dealer market and a crossing network*, Journal of Financial Economics **91** (2009), no. 3, 319–338.

[EKTZ14] I. Ekren, C. Keller, N. Touzi, and J. Zhang, *On viscosity solutions of path dependent PDEs*, The Annals of Probability **42** (2014), no. 1, 204–236. MR3161485

[Els11] J. Elstrodt, *Maß- und Integrationstheorie*, Springer-Lehrbuch, Springer Berlin Heidelberg, 2011.

[EPQ97] N. El Karoui, S. Peng, and M. C. Quenez, *Backward stochastic differential equations in finance*, Math. Finance **7** (1997), no. 1, 1–71. MR1434407

[ER09] N. El Karoui and C. Ravanelli, *Cash subadditive risk measures and interest rate ambiguity*, Math. Finance **19** (2009), no. 4, 561–590. MR2583520

[Esp10] G.-E. Espinosa, *Stochastic control methods for optimal portfolio investment*, Ph.D. Thesis, 2010.

[ET15] G.-E. Espinosa and N. Touzi, *Optimal Investment under Relative Performance Concerns*, Mathematical Finance **25** (2015), no. 2, 221–257.

[Fab10] G. Faber, *Über die Orthogonalfunktionen des Herrn Haar.*, Jahresbericht der Deutschen Mathematiker-Vereinigung **19** (1910), 104–112.

[Fag96] M.-C. Fagart, *Concurrence en contrats, anti-sélection et structure d'information*, Annales d'Économie et de Statistique **43** (1996), 1–27.

[FdR11] C. Frei and G. dos Reis, *A financial market with interacting investors: does an equilibrium exist?*, Mathematics and financial economics **4** (2011), no. 3, 161–182. MR2796281

[FH14] P. K. Friz and M. Hairer, *A Course on Rough Paths: With an Introduction to Regularity Structures*, Springer, 2014.

[FK08] D. Filipović and M. Kupper, *Equilibrium prices for monetary utility functions*, International Journal of Theoretical and Applied Finance **11** (2008), no. 3, 325–343.

[FK11] H. Föllmer and T. Knispel, *Entropic risk measures: coherence vs. convexity, model ambiguity, and robust large deviations*, Stoch. Dyn. **11** (2011), no. 2-3, 333–351. MR2836530

[Föl81] H. Föllmer, *Calcul d'Ito sans probabilités*, Séminaire de Probabilités XV 1979/80, 1981, pp. 143–150.

[FP16] S. Foley and T. J. Putniņš, *Should we be afraid of the dark? Dark trading and market quality*, Journal of Financial Economics **122** (2016), no. 3, 456–481.

[Fre14] C. Frei, *Splitting multidimensional BSDEs and finding local equilibria*, Stochastic Processes and their Applications **124** (2014), no. 8, 2654–2671.

[FRG02] M. Frittelli and E. Rosazza Gianin, *Putting order in risk measures*, Journal of Banking & Finance **26** (2002), no. 7, 1473–1486.

[FS02] H. Föllmer and A. Schied, *Convex measures of risk and trading constraints*, Finance Stoch. **6** (2002), no. 4, 429–447. MR1932379

[FV10] P. K. Friz and N. B. Victoir, *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*, Cambridge studies in advanced mathematics, vol. 120, Cambridge University Press, Cambridge, 2010.

[Góm07] J.-P. Gómez, *The impact of keeping up with the Joneses behavior on asset prices and portfolio choice*, Finance Research Letters **4** (2007), no. 2, 95–103.

[Gal94] J. Gali, *Keeping up with the Joneses: Consumption Externalities, Portfolio Choice, and Asset Prices*, Journal of Money, Credit and Banking **26** (1994), no. 1, 1–8.

[Gia06] E. R. Gianin, *Risk measures via g-expectations*, Insurance: Mathematics and Economics **39** (2006), no. 1, 19–34.

[GIP15] M. Gubinelli, P. Imkeller, and N. Perkowski, *Paracontrolled distributions and singular PDEs*, Forum of Mathematics, Pi **3** (2015), e6 (75 pages).

[GIP16] M. Gubinelli, P. Imkeller, and N. Perkowski, *A Fourier analytic approach to pathwise stochastic integration*, Electronic Journal of Probability **21** (2016), no. 2, 1–37.

[Gla04] P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Stochastic Modelling and Applied Probability, vol. 53, Springer, 2004.

[Glo94] L. Glosten, *Is the Electronic Open Limit Order Book Inevitable?*, The Journal of Finance **49** (1994), no. 4, 1127–1161.

[Gol73] B. I. Golubov, *Series with respect to the Haar system*, Journal of Soviet Mathematics **1** (1973), no. 6, 704–726.

[Gor94] R. A. Gordon, *The Integrals of Lebesgue, Denjoy, Perron, and Henstock*, Vol. 4, American Mathematical Society, 1994.

[Gre06] C. Gresse, *The Effect of Crossing-Network Trading on Dealer Market's Bid-Ask Spreads*, European Financial Management **12** (2006), no. 2, 143–160.

[GST+13] P. Gomber, S. Sagade, E. Theissen, M. C. Weber, and C. Westheide, *Competition/Fragmentation in Equities Markets: A Literature Survey*, Technical Report 35, SAFE Working Paper Series, 2013.

[GT16] E. Gobet and P. Turkedjiev, *Linear regression MDP scheme for discrete backward stochastic differential equations under general conditions*, Mathematics of Computation **85** (2016), no. 299, 1359–1391.

[Gub04] M. Gubinelli, *Controlling rough paths*, Journal of Functional Analysis **216** (2004), 86–140.

[Haa10] A. Haar, *Zur Theorie der orthogonalen Funktionensysteme. (Erste Mitteilung.)*, Mathematische Annalen **69** (1910), 331–371.

[Hil63] T. H. Hildebrandt, *Introduction to the Theory of Integration*, Pure and Applied Mathematics, vol. 13, Academic Press, 1963.

[HIM05] Y. Hu, P. Imkeller, and M. Müller, *Utility maximization in incomplete markets*, Ann. Appl. Probab. **15** (2005), no. 3, 1691–1712. MR2152241

[HM00] T. Hendershott and H. Mendelson, *Crossing Networks and Dealer Markets: Competition and Performance*, The Journal of Finance **55** (2000), no. 5, 2071–2115.

[HM07] U. Horst and M. Müller, *On the Spanning Property of Risk Bonds Priced by Equilibrium*, Mathematics of Operations Research **32** (2007), no. 4, 784–807.

[HMB11] U. Horst and S. Moreno-Bromberg, *Efficiency and equilibria in games of optimal derivative design*, Mathematics and Financial Economics **5** (2011), no. 4, 269–297.

[HN14] U. Horst and F. Naujokat, *When to Cross the Spread? Trading in Two-Sided Limit Order Books*, SIAM J. Financial Math. **5** (2014), no. 1, 278–315.

[HPDR10] U. Horst, T. Pirvu, and G. Dos Reis, *On securitization, market completion and equilibrium risk transfer*, Mathematics and Financial Economics **2** (2010), no. 4, 211–252. MR2601853

[HRAY10] R. Hinz, H. P. Rudolph, P. Antolín, and J. Yermo (eds.), *Evaluating the Financial Performance of Pension Funds*, The World Bank, 2010.

[HT16] Y. Hu and S. Tang, *Multi-dimensional backward stochastic differential equations of diagonally quadratic generators*, Stochastic Processes and their Applications **126** (2016), no. 4, 1066–1086.

[HW13] M. Hairer and H. Weber, *Rough Burgers-like equations with multiplicative noise*, Probability Theory and Related Fields **155** (2013), no. 1-2, 71–126.

[IDR10] P. Imkeller and G. Dos Reis, *Path regularity and explicit convergence rate for BSDE with truncated quadratic growth*, Stochastic Process. Appl. **120** (2010), no. 3, 348–379. MR2584898

[Imk08] P. Imkeller, *Malliavin's calculus and applications in stochastic control and finance*, IMPAN Lecture Notes, vol. 1, Institute of Mathematics, Polish Academy of Sciences, Warsaw, 2008.

[Imk15] P. Imkeller, *A Fourier approach to pathwise stochastic integration. Lectures at the 25th Jyväskylä Summer School*, 2015.

[JKL14] A. Jamneshan, M. Kupper, and P. Luo, *Multidimensional quadratic BSDEs with separated generators*, 2014. arXiv:1501.00461.

[JST06] E. Jouini, W. Schachermayer, and N. Touzi, *Law invariant risk measures have the Fatou property*, In: Advances in Mathematical Economics. Vol. 9, 2006, pp. 49–71. MR2277714

[Jul00] B. Jullien, *Participation Constraints in Adverse Selection Models*, Journal of Economic Theory **93** (2000), no. 1, 1–47.

[KK68] G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers. Definitions, Theorems, and Formulas for Reference and Review*, 2nd enl. and rev., McGraw-Hill, 1968.

[Kob00] M. Kobylanski, *Backward stochastic differential equations and partial differential equations with quadratic growth*, Ann. Probab. **28** (2000), no. 2, 558–602. MR1782267

[KP16] D. Kramkov and S. Pulido, *A system of quadratic BSDEs arising in a price impact model*, The Annals of Applied Probability **26** (2016), no. 2, 794–817.

[KS12] D. S. Kurtz and C. W. Swartz, *Theories of Integration: The Integrals of Riemann, Lebesgue, Henstock–Kurzweil, and McShane*, 2nd ed., Series in Real Analysis, vol. 13, World Scientific, 2012.

[KS15] P. Kratz and T. Schöneborn, *Portfolio liquidation in dark pools in continuous time*, Math. Finance **25** (2015), no. 3, 496—544.

[KS51] S. Kaczmarz and H. Steinhaus, *Theorie der Orthogonalreihen*, 2nd ed., Monografje Matematyczne, vol. 6, Chelsea Publishing Company, 1951.

[KTPZ15] N. Kazi-Tani, D. Possamaï, and C. Zhou, *Quadratic BSDEs with jumps: a fixed-point approach*, Electron. J. Probab. **20** (2015), no. 66, 1–28. MR3361254

[KXŽ15] C. Kardaras, H. Xing, and G. Žitković, *Incomplete stochastic equilibria with exponential utilities close to Pareto optimality*, arXiv preprint arXiv:1505.07224 (2015).

[LCL07] T. J. Lyons, M. J. Caruana, and T. Lévy, *Differential Equations Driven by Rough Paths: École d'été de Probabilités de Saint-Flour XXXIV – 2004*, Springer, 2007.

[LdRS15] A. Lionnet, G. dos Reis, and L. Szpruch, *Time discretization of FBSDE with polynomial growth drivers and reaction-diffusion PDEs*, Ann. Appl. Probab. **25** (2015), no. 5, 2563–2625. MR3375884

[Lea01] S. Leader, *The Kurzweil-Henstock Integral and Its Differential: A Unified Theory of Integration on $\mathbb{R}$ and $\mathbb{R}^n$*, CRC Press, 2001.

[LQ02] T. Lyons and Z. Qian, *System Control and Rough Paths*, Oxford University Press, 2002.

[McL80] R. M. McLeod, *The Generalized Riemann Integral*, 1st ed., The Carus Mathematical Monographs, vol. 20, Mathematical Association of America, 1980.

[MPR17] T. Mastrolia, D. Possamaï, and A. Réveillac, *On the Malliavin differentiability of BSDEs*, Ann. Inst. H. Poincaré Probab. Statist. **53** (2017), no. 1, 464–492.

[MR78] M. Mussa and S. Rosen, *Monopoly and Product Quality*, Journal of Economic Theory **18** (1978), 301–317.

[MY10] J. Ma and S. Yao, *On Quadratic g-Evaluations/Expectations and Related Analysis*, Stochastic Analysis and Applications **28** (2010), no. 4, 711–734. MR2739601

[Mye91] R. B. Myerson, *Game Theory: Analysis of Conflict*, Harvard University Press, United States of America, 1991.

[NØ06] R. Næs and B. A. Ødegaard, *Equity trading by institutional investors: To cross or not to cross?*, Journal of Financial Markets **9** (2006), no. 2, 79–99.

[Neg60] T. Negishi, *Welfare economics and existence of an equilibrium for a competitive economy*, Metroeconomica **12** (1960), no. 2-3, 92–97.

[Nie97] O. A. Nielsen, *An Introduction to Integration and Measure Theory*, Wiley-Interscience, 1997.

[Nou12] I. Nourdin, *Selected Aspects of Fractional Brownian Motion*, Bocconi & Springer Series, Springer-Verlag, Mailand, 2012.

[NR14] M. Nimalendran and S. Ray, *Informational linkages between dark and lit trading venues*, Journal of Financial Markets **17** (2014), 230–261.

[Nua06] D. Nualart, *The Malliavin Calculus and Related Topics*, 2nd ed., Probability and its Applications (New York), Springer-Verlag, Berlin, 2006.

[Øk97] B. Øksendal, *An Introduction to Malliavin Calculus with Applications to Economics. Lecture notes*, 1997.

[Ori12] N. Oriol, *Fragmentation of Order Flows and Revision of MiFID: Lessons from Industrial Economics*, Revue d'économie industrielle **3** (2012), 49–76.

[Pag92] F. H. Page Jr., *Mechanism design for general screening problems with moral hazard*, Economic Theory **2** (1992), no. 2, 265–281.

[Ped15] A. Pedraza, *Strategic Interactions and Portfolio Choice in Money Management: Theory and Evidence*, Journal of Money, Credit and Banking **47** (2015), no. 8, 1531–1569.

[Pen97] S. Peng, *Backward SDE and related g-expectation*, In: N. El Karoui and L. Mazliak (eds.), Backward stochastic differential equations, 1997, pp. 141–159. MR1752680

[Pfe93] W. F. Pfeffer, *The Riemann approach to integration: Local geometric theory*, Vol. 109, Cambridge University Press, 1993.

[PP16] N. Perkowski and D. J. Prömel, *Pathwise stochastic integrals for model free finance*, Bernoulli **22** (2016), no. 4, 2486–2520.

[PS03] C. A. Parlour and D. J. Seppi, *Liquidity-Based Competition for Order Flow*, The Review of Financial Studies **16** (2003), no. 2, 301–343.

[PSS$^+$08] J. Pouyet, B. Salanié, F. Salanié, et al., *On competitive equilibria with asymmetric information*, The B.E. Journal of Theoretical Economics **8** (2008), no. 1, 1–16.

[Rüs13] L. Rüschendorf, *Mathematical Risk analysis. Dependence, Risk Bounds, Optimal Allocations and Portfolios*, Springer Series in Operations Research and Financial Engineering, Springer, 2013.

[REK00] R. Rouge and N. El Karoui, *Pricing Via Utility Maximization and Entropy*, Mathematical Finance **10** (2000), no. 2, 259–276.

[RH73] H. E. Ryder Jr. and G. M. Heal, *Optimal Growth with Intertemporally Dependent Preferences*, The Review of Economic Studies **40** (1973), no. 1, 1–31.

[Roc85] J.-C. Rochet, *The taxation principle and multi–time Hamilton-Jacobi equations*, Journal of Mathematical Economics **14** (1985), 113–128.

[Roy93] B. Roynette, *Mouvement brownien et espaces de Besov*, Stochastics and Stochastic Reports **43** (1993), no. 3-4, 221–260.

[Rud76] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed., McGraw-Hill New York, 1976.

[Sch17] D. C. Schwarz, *Market completion with derivative securities*, Finance and Stochastics **21** (2017), no. 1, 263–284.

[Sch27] J. Schauder, *Zur Theorie stetiger Abbildungen in Funktionalräumen*, Mathematische Zeitschrift **26** (1927), no. 1, 47–65.

[Sch96] E. Schechter, *Handbook of analysis and its foundations*, Academic Press, 1996.

[SUV07] J. L. Solé, F. Utzet, and J. Vives, *Canonical Lévy process and Malliavin calculus*, Stochastic Processes and their Applications **117** (2007), no. 2, 165–187.

[Swa01] C. Swartz, *Introduction to gauge integrals*, World Scientific, 2001.

[Tar55] A. Tarski, *A Lattice-Theoretical Fixpoint Theorem and its Applications*, Pacific J. Math. **5** (1955), no. 2, 285–309.

[Tev08] R. Tevzadze, *Solvability of backward stochastic differential equations with quadratic growth*, Stochastic Processes and their Applications **118** (2008), no. 3, 503–515. MR2389055

[Tou13] N. Touzi, *Optimal Stochastic Control, Stochastic Target Problems, and Backward SDE*, Fields Institute Monographs, vol. 29, Springer, New York, 2013. With Chapter 13 by Angès Tourin. MR2976505

[Ul'64] P. L. Ul'yanov, *On Haar series*, Matematicheskii Sbornik **63(105)** (1964), no. 3, 356–391.

[vBC00] M. van Brunt and B. Carter, *The Lebesgue-Stieltjes Integral. A Practical Introduction*, Springer, 2000.

[Vol16] I. Voloshchenko, *On pathwise functional Itô calculus and its applications to mathematical finance*, Ph.D. Thesis, 2016.

[Wes51] J. D. Weston, *Inequalities for Riemann-Stieltjes Integrals*, Mathematische Zeitschrift **54** (1951), no. 3, 272–274.

[XZ10] C. Xiouros and F. Zapatero, *The Representative Agent of an Economy with External Habit Formation and Heterogeneous Risk Aversion*, Review of Financial Studies **23** (2010), no. 8, 3017–3047.

[XŽ16] H. Xing and G. Žitković, *A class of globally solvable Markovian quadratic BSDE systems and applications*, arXiv preprint arXiv:1603.00217 (2016).

[Ye16] L. Ye, *Understanding the impacts of dark pools on price discovery*, arXiv preprint arXiv:1612.08486 (2016).

[You36] L. C. Young, *An inequality of the Hölder type, connected with Stieltjes integration*, Acta Mathematica **67** (1936), no. 1, 251–282.

[Zhu14] H. Zhu, *Do Dark Pools Harm Price Discovery?*, The Review of Financial Studies **27** (2014), no. 3, 747–789.

Acknowledgments

First, I would like to thank my supervisor, Ulrich Horst, for his support and his patience throughout these past years and also for his honesty and understanding.

I am also very thankful to Peter Imkeller, who did not only give brilliant lectures, but who also offered me a very interesting topic to work on.

Moreover, I want to thank Gonçalo dos Reis, Arnaud Lionnet and Santiago Moreno-Bromberg for the fruitful discussions and good advice.

I also thank all my colleagues from Humboldt-Universität zu Berlin, in particular Sabine Bergmann and Andrea Fiebig, whose help with bureaucracy is priceless.

Last but not least I also want to thank my family and friends for their constant support and encouragement.

Selbständigkeitserklärung

Ich erkläre, dass ich die Dissertation selbständig und nur unter Verwendung der von mir gemäß § 7 Abs. 3 der Promotionsordnung der Mathematisch-Naturwissenschaftlichen Fakultät, veröffentlicht im Amtlichen Mitteilungsblatt der Humboldt-Universität zu Berlin Nr. 126/2014 am 18.11.2014 angegebenen Hilfsmittel angefertigt habe.

Berlin, 27.03.2017